SEMIVARIOGRAM ESTIMATION: ASYMPTOTIC THEORY AND APPLICATIONS

by

Brent Kerby

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

The University of Utah

May 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Brent Kerby**

has been approved by the following supervisory committee members:

Lajos Horvath ,	Chair(s)	**17 Feb 2016** Date Approved
Davar Khoshnevisan ,	Member	**17 Feb 2016** Date Approved
Stewart Ethier ,	Member	**17 Feb 2016** Date Approved
Yekaterina Epshteyn ,	Member	**17 Feb 2016** Date Approved
Ron Reeder ,	Member	**17 Feb 2016** Date Approved

by **Peter Trapa** , Chair/Dean of

the Department/College/School of **Mathematics**

and by **David B. Kieda** , Dean of The Graduate School.

ABSTRACT

The semivariogram is a function characterizing the second-order dependence structure of an intrinsically stationary random field; its estimation has applications in spatial statistics, particularly in the construction of optimal predictors of the random field at unobserved locations. In this work, we establish conditions under which the empirical isotropic semivariogram converges to the semivariogram uniformly on compact sets. In preparation for these results, we also establish sufficient conditions for stationary Gaussian random fields to be ρ^*-mixing, in terms of the spectral density. We also introduce two new applications of semivariogram estimation: a method for digital image compression, and a refinement of the Moran's I test for spatial autocorrelation.

For my parents, Tom and June.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOTATION AND SYMBOLS

$a_k \ll b_k$	a_k asymptotically grows no faster than b_k: $\limsup_{k\to\infty} \frac{a_k}{b_k} < \infty$.
$a_k \asymp b_k$	a_k asymptotically grows at the same rate as b_k: both $a_k \ll b_k$ and $b_k \ll a_k$.
$B_A(r)$	$= \{x \in \mathbb{R}^n : d(A, \{x\}) < r\}$, open neighborhood of radius r of the subset $A \subseteq R^n$
$C(\boldsymbol{h})$	$= \mathrm{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$, covariance function of random field $Z(\boldsymbol{s})$
$C^o(\|\boldsymbol{h}\|)$	$= \mathrm{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$, isotropic covariance function
$\hat{C}(\boldsymbol{h})$	empirical covariance function; see (1.10).
$\hat{C}^o(h)$	empirical isotropic covariance function; see (1.11).
$C_c(X)$	the set of real compactly-supported continuous functions on a topological space X
$d(A, B)$	$= \inf\{\|x-y\| : x \in A, y \in B\}$, distance between subsets $A, B \subseteq \mathbb{R}^n$.
$\ker L$	$\{x : L(x) = 0\}$, kernel of a linear transformation L
L^p	the space of a.s. equivalence classes of random variables X with $E(\|X\|^p) < \infty$
$\mathrm{sgn}\, x$	sign of a real number: 1 if $x > 0$, 0 if $x = 0$, and -1 if $x < 0$
$\boldsymbol{e}_i$	ith standard basis vector of $\mathbb{R}^n$: $(0, \ldots, 1, \ldots, 0)$, 1 in ith position, 0 elsewhere
$\gamma(\boldsymbol{h})$	$= \frac{1}{2}E\left[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2\right]$, semivariogram of the random field $Z(\boldsymbol{s})$
$\gamma^o(\|\boldsymbol{h}\|)$	$= \frac{1}{2}E\left[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2\right]$, isotropic semivariogram
$\hat{\gamma}(\boldsymbol{h})$	empirical semivariogram; see (1.12).
$\hat{\gamma}^o(h)$	empirical isotropic semivariogram; see (1.13).
$\hat{\gamma}_\delta(h)$	empirical isotropic semivariogram with bandwidth δ; see (1.14).
$\lambda(A)$	Lebesgue measure of the subset $A \subseteq \mathbb{R}^n$
$\Phi(x)$	the standard Gaussian cumulative distribution function
$(\Omega, \mathcal{F}, P)$	probability space Ω with σ-algebra of events $\mathcal{F}$ and probability measure P
$\|A\|$	cardinality of a finite set A
$\mathbb{R}^{n\times n}$	the set of $n \times n$ matrices with real entries

CHAPTER 1

INTRODUCTION

Spatial statistics is concerned with quantities exhibiting random variation across space. The initial development of the subject was driven by problems in mining engineering [25], with applications now extending to diverse areas including agriculture, criminology, econometrics, ecology, and astronomy. In spatial statistics, the classical assumption of independent identically distributed observations is abandoned in favor of an assumption that observations exhibit dependence based on their spatial proximity. Spatial statistics borrows many techniques from the study of time series and stochastic processes, extending one-dimensional methods to multidimensional settings. Background on the basic concepts and methods of spatial statistics may be found in [16], [36], and [9]. Proofs of our results, along with additional technical background and details, may be found at the end of the chapter.

1.1 Spatial statistical modeling

Let $Z(\boldsymbol{s})$ represent a random, real quantity varying across locations $\boldsymbol{s}$ ranging over some region $D \subseteq \mathbb{R}^n$. For example, $Z(\boldsymbol{s})$ might represent the concentration of gold at location $\boldsymbol{s}$ of a prospective mining site. Note that even though $Z(\boldsymbol{s})$ may only be observed at finitely many sites $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$, we nevertheless consider $Z(\boldsymbol{s})$ to be a random field defined on all of D. For many types of spatially varying data, measurements of $Z(\boldsymbol{s})$ at two nearby locations tend to differ less from one another than measurements at two widely separated locations. To account for this, a classical statistical approach would postulate that the mean $E(Z(\boldsymbol{s}))$ is a non-constant function of $\boldsymbol{s}$, motivating a linear model

$$Z(\boldsymbol{s}_i) = \mu(\boldsymbol{s}_i) + \epsilon(\boldsymbol{s}_i)$$
$$\mu(\boldsymbol{s}_i) = \boldsymbol{X}(\boldsymbol{s}_i)^T \boldsymbol{\beta}$$

where the errors $\epsilon(s_i)$ are assumed to be independent and identically distributed with mean zero, and where the components of $\boldsymbol{X}(\boldsymbol{s})$ are for example a suitable set of polynomials in $\boldsymbol{s}$. The parameter vector $\boldsymbol{\beta}$ may then be estimated by the method of least squares. This

approach is known as *trend surface analysis*. One of its chief deficiencies is that in order to adequately model the small-scale spatial variation, $\boldsymbol{\beta}$ may be required to have very large dimension.

In contrast, a spatial statistical approach, in its most basic form, postulates a constant mean $\mu = E(Z(\boldsymbol{s}))$ and a covariance depending on the displacement between the two locations; i.e., we assume that the function $C(\boldsymbol{h}) = \mathrm{Cov}(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h}))$ is well-defined, not depending on the choice of $\boldsymbol{s}$. In this case, we say that the random field $Z(\boldsymbol{s})$ is *second-order stationary*. Often some form of parametric model is imposed on the covariance function, and an adequate model can commonly be obtained with only a small number of parameters. Two examples of parametric models for covariance functions, which we will make use of later on, are the exponential and Gaussian models (see Theorem 1.12 for a proof that these are valid covariance functions):

$$\text{Exponential: } C(\boldsymbol{h}) = \sigma^2 e^{-\|\boldsymbol{h}\|/h_0}$$
$$\text{Gaussian: } C(\boldsymbol{h}) = \sigma^2 e^{-\frac{1}{2}(\|\boldsymbol{h}\|/h_0)^2}$$

Given any second-order stationary random field, the covariance function has the property of even symmetry; i.e., $C(\boldsymbol{h}) = C(-\boldsymbol{h})$ for all $\boldsymbol{h} \in \mathbb{R}^n$. If, furthermore, $C(\boldsymbol{h})$ depends on $\boldsymbol{h}$ only through its length $\|\boldsymbol{h}\|$, as in the examples above, then we say that the random field is *second-order isotropic*, and we define the *isotropic covariance function* $C^o(h)$ by $C^o(\|\boldsymbol{h}\|) = C(\boldsymbol{h})$.

The class of possible covariance functions $C(\boldsymbol{h})$ is quite large. An even function $C(\boldsymbol{h})$ is the covariance function of some second-order stationary random field if and only if $C(\boldsymbol{h})$ is *positive semidefinite*, i.e., if and only if

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i - \boldsymbol{s}_j) \geq 0$$

for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^n$ (see Theorem 1.7). These functions are described explicitly by Bochner's theorem [35, §1.4.3]:

Theorem 1.1. *An even, continuous function $C : \mathbb{R}^n \to \mathbb{R}$ is positive semidefinite if and only if there exists a finite, positive Borel measure μ on $\mathbb{R}^n$ such that*

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^n} e^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{h}}\, d\mu(\boldsymbol{\omega}), \tag{1.1}$$

and in this case, μ is uniquely determined by C.

In this case, we call μ the *spectral measure* associated with C. If μ is absolutely continuous with respect to Lebesgue measure λ, with $d\mu = f\ d\lambda$ for some function $f(\omega)$ on $\mathbb{R}^n$, then we call f (which is uniquely determined almost everywhere) the *spectral density*. The fact that C is real valued, together uniqueness of μ, implies that μ must be symmetric, in the sense that $\mu(A) = \mu(-A)$ for every Borel measurable set $A \subseteq \mathbb{R}^n$. Therefore, (1.1) may be put into an alternative form:

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^n} \cos(\boldsymbol{\omega} \cdot \boldsymbol{h})\ d\mu(\boldsymbol{\omega}), \tag{1.2}$$

1.2 Semivariogram of random field

Alongside the covariance function, an alternative approach to describing the dependence structure of a random field is the *semivariogram*, introduced by Mathéron [31]:

$$\gamma(\boldsymbol{h}) = \frac{1}{2} E\left[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2\right]. \tag{1.3}$$

A random field for which (1.3) is well-defined (i.e., for which the expectation exists and does not depend on $\boldsymbol{s}$) is said to be *intrinsically stationary*. If $Z(\boldsymbol{s})$ is second-order stationary, then $Z(\boldsymbol{s})$ is also intrinsically stationary, for in this case, we have

$$\begin{aligned}
\frac{1}{2} E\left[(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2\right] &= \frac{1}{2} \operatorname{Var}(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})) \\
&= \frac{1}{2} \left[\operatorname{Var}(Z(\boldsymbol{s}+\boldsymbol{h})) + \operatorname{Var}(Z(\boldsymbol{s})) - 2\operatorname{Cov}(Z(\boldsymbol{s}+\boldsymbol{h}), Z(\boldsymbol{s}))\right] \\
&= \frac{1}{2}[C(\boldsymbol{0}) + C(\boldsymbol{0}) - 2C(\boldsymbol{h})] \\
&= C(\boldsymbol{0}) - C(\boldsymbol{h}),
\end{aligned}$$

where $C(\boldsymbol{h})$ is the covariance function of Z. Thus, in this case, the semivariogram γ is completely determined by the covariance function C through the simple formula

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h}). \tag{1.4}$$

However, there are random fields which are intrinsically stationary but not second-order stationary. For example, Lévy's Brownian motion is an intrinsically stationary Gaussian random field on $\mathbb{R}^n$, given by the semivariogram $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}\|$ with $Z(\boldsymbol{0}) = 0$, but it has non-constant variance $\operatorname{Var}(\boldsymbol{s}) = \|\boldsymbol{s}\|$ and hence is not second-order stationary (see Theorems 1.10 and 1.13). In a case such as this, where the semivariogram $\gamma(\boldsymbol{h})$ depends on $\boldsymbol{h}$ only through its length $\|\boldsymbol{h}\|$, we define the *isotropic semivariogram* to be the function $\gamma^o(h)$ given by $\gamma^o(\|\boldsymbol{h}\|) = \gamma(\boldsymbol{h})$.

An even function $\gamma(\boldsymbol{h})$ is the semivariogram of some intrinsically stationary random field if and only if $\gamma(\boldsymbol{h})$ is *conditionally negative semidefinite*, i.e., if and only if

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \leq 0$$

for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^n$ such that $\sum_{i=1}^{k} a_i = 0$ (see Theorem 1.10 and also [16, p. 36]). The conditions of positive semidefiniteness and conditionally negative semidefiniteness are related as follows [21, Lemma 1]:

Theorem 1.2. *An even function $\gamma(\boldsymbol{h})$ on $\boldsymbol{h} \in \mathbb{R}^n$ is conditionally negative semidefinite if and only if $e^{-a\gamma(\boldsymbol{h})}$ is positive semidefinite for all $a > 0$.*

Given a second-order stationary random field, it is common to assume that $\lim_{h\to\infty} C(h) = 0$. In this case, the semivariogram $\gamma(h)$ approaches a limit, called the *sill* of the semivariogram:

$$\lim_{h\to\infty} \gamma(h) = \lim_{h\to\infty} [C(0) - C(h)] = C(0)$$

The semivariogram is said to have *finite range* if it reaches its sill after a finite distance, i.e., if $C(h)$ is exactly zero for sufficiently large h.

On a historical note, we mention that Mathéron originally defined the semivariogram not as an expected value of a random variable, but as a spatial integral of a deterministic function [30]:

$$\gamma(\boldsymbol{h}) = \frac{1}{2\lambda(V)} \iiint_V (z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}))^2 \, d\boldsymbol{s} \tag{1.5}$$

While Mathéron eventually turned to probabilistic methods for their practical utility in enabling statistical inference, he expressed philosophical reservations about interpreting spatial data as observations of a random field when only one realization of the field is available for observation; he believed that the non-probabilistic definition of $\gamma(\boldsymbol{h})$ was an alternative approach leading to equivalent results [31, p. 6]. He postulated that the integral in (1.5) would give the same value regardless of the choice of the region V, which may be regarded as a deterministic version of an assumption of stationarity. However, it turns out that this assumption is too strong, in that any function $z(\boldsymbol{s})$ satisfying such an assumption degenerates into being equal almost everywhere to a linear function plus a constant, as we show in Theorem 1.17. Although in retrospect it appears that such a non-probabilistic approach cannot provide a suitable basis for spatial statistics, it is worth noting that under certain conditions, an integral of the form (1.5) does make its appearance as a limit of

semivariogram estimators based on an increasing number of sample points on a bounded region V (see Lahiri [26, Theorem 2]).

1.3 Prediction on random fields

For many applications, the main goal of modeling a spatial field is to enable prediction at unobserved locations. For example, a mining engineer, having measured the concentration of an ore at a sample of locations, may wish to predict the concentration throughout the deposit. In geostatistics, the process of forming such predictions is known as *kriging*, in honor of the mining engineer Danie Krige who pioneered the field [24]. In its most basic form, kriging involves constructing a best linear unbiased predictor. More precisely, given an intrinsically stationary random field $Z(\boldsymbol{s})$ with unknown mean μ but known semivariogram $\gamma(\boldsymbol{h})$, if we observe Z at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ and wish to predict Z at $\boldsymbol{s}_0$, then the kriging predictor $\hat{Z}(\boldsymbol{s}_0)$ is defined as the linear combination

$$\hat{Z}(\boldsymbol{s}_0) = \sum_{i=1}^{m} \lambda_i Z(\boldsymbol{s}_i) \tag{1.6}$$

where the coefficients $\lambda_1, \ldots, \lambda_m$ are chosen to ensure that the error $Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0)$ has mean zero and minimum possible variance. Since $E(Z(\boldsymbol{s}_0)) = \mu$ while $E(\hat{Z}(\boldsymbol{s}_0)) = \sum_{i=1}^{m} \lambda_i \mu$, the condition that $Z(\boldsymbol{s}_0) - \hat{Z}(\boldsymbol{s}_0)$ has mean zero (for all possible values of μ) is equivalent to the condition

$$\sum_{i=1}^{m} \lambda_i = 1 \tag{1.7}$$

Assuming this condition holds, we may calculate the variance of the error:

$$\begin{aligned}
E\left[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)\right]^2 &= E\left[\sum_{i=1}^{m} \lambda_i (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_0))\right]^2 \\
&= E\left[\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_0))(Z(\boldsymbol{s}_j) - Z(\boldsymbol{s}_0))\right] \\
&= E\left[Z(\boldsymbol{s}_0)^2 - 2\sum_{i=1}^{m} \lambda_i Z(\boldsymbol{s}_0) Z(\boldsymbol{s}_i) + \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j Z(\boldsymbol{s}_i) Z(\boldsymbol{s}_j)\right] \\
&= E\left[\sum_{i=1}^{m} \lambda_i (Z(\boldsymbol{s}_0) - Z(\boldsymbol{s}_i))^2 - \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j \frac{(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2}{2}\right] \\
&= 2\sum_{i=1}^{m} \lambda_i \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)
\end{aligned} \tag{1.8}$$

To minimize this quantity subject to the constraint $\sum_{i=1}^m \lambda_i = 1$, one may introduce a Lagrange multiplier λ_0 and obtain equations

$$\gamma(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{j=1}^{m} \lambda_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = \lambda_0 \tag{1.9}$$

for all $i = 1, \ldots, m$. Letting $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$, and defining a matrix Γ and vector $\boldsymbol{v}$ by

$$\Gamma = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_1) & \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2) & \cdots & \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_m) \\ 1 & \gamma(\boldsymbol{s}_2 - \boldsymbol{s}_1) & \gamma(\boldsymbol{s}_2 - \boldsymbol{s}_2) & \cdots & \gamma(\boldsymbol{s}_2 - \boldsymbol{s}_m) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \gamma(\boldsymbol{s}_m - \boldsymbol{s}_1) & \gamma(\boldsymbol{s}_m - \boldsymbol{s}_2) & \cdots & \gamma(\boldsymbol{s}_m - \boldsymbol{s}_m) \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} 1 \\ \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_2) \\ \vdots \\ \gamma(\boldsymbol{s}_0 - \boldsymbol{s}_m) \end{pmatrix}$$

equations (1.7) and (1.9) determine a linear system

$$\Gamma \boldsymbol{\lambda} = \boldsymbol{v}$$

Therefore, if Γ is nonsingular, then the kriging predictor is defined by (1.6) with $\boldsymbol{\lambda} = \Gamma^{-1}\boldsymbol{v}$. This is known as the *ordinary kriging predictor*. The variance of the predictor is given by (1.8), which enables the construction of prediction intervals if the random field is assumed to be Gaussian. Other types of kriging predictors are obtained in a similar manner if we vary the assumptions of the model: if we assume that the mean μ is known, then the result is known as *simple kriging*, while if we assume that $Z(\boldsymbol{s})$ is given by a linear model

$$Z(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^T \boldsymbol{\beta} + \epsilon(\boldsymbol{s})$$

with non-constant mean $\mu(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^T\boldsymbol{\beta}$ and intrinsically stationary errors $\epsilon(\boldsymbol{s})$, then the result is known as *universal kriging*; for these and other variations, along with techniques for dealing with the numerical and computational difficulty of inverting Γ, see standard texts on the subject [16, 36, 9].

If Z is second-order stationary, then by applying (1.4) to (1.9) and substituting $-\lambda_0$ for λ_0, we have

$$C(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{j=1}^{m} \lambda_j C(\boldsymbol{s}_i - \boldsymbol{s}_j) = \lambda_0$$

so that we may also express $\boldsymbol{\lambda}$ in terms of the covariance function, as $\boldsymbol{\lambda} = \Sigma^{-1}\boldsymbol{u}$ where

$$\Sigma = \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & C(\boldsymbol{s}_1 - \boldsymbol{s}_1) & C(\boldsymbol{s}_1 - \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_1 - \boldsymbol{s}_m) \\ 1 & C(\boldsymbol{s}_2 - \boldsymbol{s}_1) & C(\boldsymbol{s}_2 - \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_2 - \boldsymbol{s}_m) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & C(\boldsymbol{s}_m - \boldsymbol{s}_1) & C(\boldsymbol{s}_m - \boldsymbol{s}_2) & \cdots & C(\boldsymbol{s}_m - \boldsymbol{s}_m) \end{pmatrix}, \quad \boldsymbol{u} = \begin{pmatrix} 1 \\ C(\boldsymbol{s}_0 - \boldsymbol{s}_1) \\ C(\boldsymbol{s}_0 - \boldsymbol{s}_2) \\ \vdots \\ C(\boldsymbol{s}_0 - \boldsymbol{s}_m) \end{pmatrix}$$

For large sample sizes, numerical and computational difficulties arise when attempting to calculate Σ^{-1}; many approaches to circumventing these difficulties are discussed in the literature [8, 2, 19, 32].

1.4 Estimation of semivariogram

In the previous section, we assumed that the semivariogram $\gamma(\boldsymbol{h})$ or covariance function $C(\boldsymbol{h})$ was known. However, typically in practice this is not the case: instead we must estimate $\gamma(\boldsymbol{h})$ or $C(\boldsymbol{h})$ based on observations at a sample of locations. If the observation locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ are regularly spaced on a grid, then for suitable vectors $\boldsymbol{h}$ there will be many locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ separated by a vector $\boldsymbol{h}$, i.e., such that $\boldsymbol{s}_j - \boldsymbol{s}_i = \boldsymbol{h}$. In this case, given a second-order stationary random field $Z(\boldsymbol{s})$, we may calculate its *empirical covariance function*,

$$\hat{C}(\boldsymbol{h}) = \frac{1}{N(\boldsymbol{h})} \sum (Z(\boldsymbol{s}_i) - \overline{Z})(Z(\boldsymbol{s}_j) - \overline{Z}) \tag{1.10}$$

where the sum ranges over all pairs of observation locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\boldsymbol{s}_j - \boldsymbol{s}_i = \boldsymbol{h}$, $N(\boldsymbol{h})$ is the number of such pairs, and $\overline{Z}$ is the sample mean taken over all observations. The estimator $\hat{C}(\boldsymbol{h})$ satisfies the symmetry property $\hat{C}(\boldsymbol{h}) = \hat{C}(-\boldsymbol{h})$. Of course, $\hat{C}$ is only a partially defined function on $\mathbb{R}^n$, since for most vectors $\boldsymbol{h}$ there will be no pairs of observation locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ with $\boldsymbol{s}_j - \boldsymbol{s}_i = \boldsymbol{h}$.

If the random field $Z(\boldsymbol{s})$ is assumed to be second-order isotropic, then it is natural to instead consider the *empirical isotropic covariance function*

$$\hat{C}^o(h) = \frac{1}{N^o(h)} \sum (Z(\boldsymbol{s}_i) - \overline{Z})(Z(\boldsymbol{s}_j) - \overline{Z}) \tag{1.11}$$

where the sum ranges over all pairs of observation locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\|\boldsymbol{s}_j - \boldsymbol{s}_i\| = h$, $N^o(h)$ is the number of such pairs, and $\overline{Z}$ is the sample mean taken over all observations.

The semivariogram can be estimated in a natural way, analagously to (1.10):

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum (Z(\boldsymbol{s}_j) - Z(\boldsymbol{s}_i))^2 \tag{1.12}$$

This is known as the *empirical semivariogram*. Likewise, the *empirical isotropic semivariogram* is defined as

$$\hat{\gamma}^o(h) = \frac{1}{2N^o(h)} \sum (Z(\boldsymbol{s}_j) - Z(\boldsymbol{s}_i))^2 \tag{1.13}$$

where the sum ranges over all pairs of observation locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ such that $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| = h$, and again $N^o(h)$ is the number of such pairs.

In (1.4) we saw that for a second-order stationary random field, the semivariogram may be described in terms of the covariance function:

$$\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$$

However, it is worth noting that this relationship does not hold for the corresponding estimators: in general, $\hat{\gamma}(\boldsymbol{h}) \neq \hat{C}(\boldsymbol{0}) - \hat{C}(\boldsymbol{h})$. In fact we may calculate

$$\begin{aligned}
\hat{\gamma}(\boldsymbol{h}) &= \frac{1}{2N(\boldsymbol{h})} \sum (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2 \\
&= \frac{1}{2N(\boldsymbol{h})} \sum ((Z(\boldsymbol{s}_i) - \overline{Z}) - (Z(\boldsymbol{s}_j) - \overline{Z}))^2 \\
&= \frac{1}{2N(\boldsymbol{h})} \sum ((Z(\boldsymbol{s}_i) - \overline{Z})^2 + (Z(\boldsymbol{s}_j) - \overline{Z})^2 - 2(Z(\boldsymbol{s}_i) - \overline{Z})(Z(\boldsymbol{s}_j) - \overline{Z})) \\
&= \frac{1}{2N(\boldsymbol{h})} \sum ((Z(\boldsymbol{s}_i) - \overline{Z})^2 + (Z(\boldsymbol{s}_j) - \overline{Z})^2) - \hat{C}(\boldsymbol{h})
\end{aligned}$$

where each sum is taken over all pairs of points $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ with $\boldsymbol{s}_j - \boldsymbol{s}_i = \boldsymbol{h}$. Assuming $\boldsymbol{h} \neq 0$, we now observe that $\hat{\gamma}(\boldsymbol{h}) = \hat{C}(\boldsymbol{0}) - \hat{C}(\boldsymbol{h})$ if and only if the number of points $\boldsymbol{s}_j$ with $\boldsymbol{s}_j - \boldsymbol{s}_i = \pm\boldsymbol{h}$ does not depend on i, which is a specialized condition that does not hold in general.

Given a random field with semivariogram γ, it is clear that $\hat{\gamma}(\boldsymbol{h})$ is an unbiased estimator of $\gamma(\boldsymbol{h})$ for all $\boldsymbol{h}$. Likewise, given a random field with isotropic semivariogram γ^o, $\hat{\gamma}^o(h)$ is an unbiased estimator of $\gamma^o(h)$. The same, however, cannot be said for the estimators $\hat{C}(\boldsymbol{h})$ and $\hat{C}^o(h)$, which in general are biased. For example, in the simple case where we observe a second-order stationary random field at only two locations $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, then letting $\boldsymbol{h} = \boldsymbol{s}_2 - \boldsymbol{s}_1$ we have

$$\begin{aligned}
E[\hat{C}(\boldsymbol{0})] &= \frac{1}{2} E[(Z(\boldsymbol{s}_1) - \overline{Z})^2 + (Z(\boldsymbol{s}_2) - \overline{Z})^2] \\
&= E\left[\left(\frac{1}{2}Z(\boldsymbol{s}_1) - \frac{1}{2}Z(\boldsymbol{s}_2)\right)^2\right] \\
&= \frac{1}{2}\gamma(\boldsymbol{h}) \\
&= \frac{1}{2}C(\boldsymbol{0}) - \frac{1}{2}C(\boldsymbol{h}),
\end{aligned}$$

so here $\hat{C}(\boldsymbol{0})$ is an unbiased estimator for $C(\boldsymbol{0})$ if and only if $C(\boldsymbol{0}) = -C(\boldsymbol{h})$, which is not generally the case.

In the case where locations are irregularly spaced, the estimators (1.12) and (1.13) are useless since for any $\boldsymbol{h}$ typically at most one pair of points will be separated by $\boldsymbol{h}$ or

by a distance of $\|\boldsymbol{h}\|$. In the isotropic case, one approach is to modify the estimator by introducing a bandwidth $\delta > 0$ and defining

$$\hat{\gamma}_\delta(h) = \frac{1}{2N_\delta(h)} \sum (Z(\boldsymbol{s}_j) - Z(\boldsymbol{s}_i))^2 \tag{1.14}$$

where the sum ranges over all pairs of observation locations $\boldsymbol{s}_j$ and $\boldsymbol{s}_i$ such that $h - \delta < \|\boldsymbol{s}_j - \boldsymbol{s}_i\| < h + \delta$. The behavior of the estimator depends greatly on the choice of δ: choosing δ too small leads to a large variance due to there being few terms in the sum, while choosing δ too large leads to a large bias due to the inclusion of many terms where $\|\boldsymbol{s}_j - \boldsymbol{s}_i\|$ is not close to h.

The estimator in (1.14) may be generalized to a kernel estimator, also known as a Nadaraya-Watson semivariogram estimator, considered by Hall et al. [17] and Garcia-Soidán et al. [14],

$$\hat{\gamma}_\delta^K(h) = \frac{\sum K\left(\frac{h - \|\boldsymbol{s}_j - \boldsymbol{s}_i\|}{\delta}\right) (Z(\boldsymbol{s}_j) - Z(\boldsymbol{s}_i))^2}{2 \sum K\left(\frac{h - \|\boldsymbol{s}_j - \boldsymbol{s}_i\|}{\delta}\right)} \tag{1.15}$$

where K is a function, and the sum is taken over all pairs of observations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. For example, (1.14) is recovered by taking K to be the indicator function of the open interval (0,1), while, if instead K is chosen to be a continuous, decreasing function supported in (0,1), then we may regard (1.15) as a smoothed version of (1.14).

1.5 Asymptotics of semivariogram estimation

When forming kriging predictors, if the semivariogram is replaced by an estimated semivariogram, then we no longer have the assurance that the predictors are optimal, and the variance formula (1.8) is generally no longer correct. Nevertheless, for large samples such a procedure may be justified if it can be shown that the semivariogram estimator is consistent. Much of our work is focused on establishing such consistency results (see Chapter 3).

However, even in cases where they are consistent, the empirical semivariogram and its variants discussed in the previous section are not suitable for use directly in kriging. The main difficulties are that these estimated semivariograms are undefined for large $\boldsymbol{h}$, and they may not satisfy the property of conditional negative definiteness. For these reasons, it is common to use least squares to fit a valid semivariogram model to the empirical semivariogram. Typically, this is done using a parametric model, although Shapiro and Botha [38] and Hall et al. [17, 18] discuss nonparametric methods based on inverting an estimate of the spectral density. Lahiri, Lee, and Cressie have shown that under

mild assumptions, if a semivariogram estimator such as the empirical semivariogram is consistent, then the estimators of semivariogram model parameters derived from it by the least squares method will also be consistent [27]. For this result to apply, therefore, it is first necessary to ensure that the empirical semivariogram is consistent. We also mention that likelihood estimation, including maximum likelihood and restricted maximum likelihood (REML) methods, can be used to fit a parametric semivariogram model directly to the data, sidestepping the need to compute the empirical semivariogram altogether; in this case, however, the known asymptotic results apply only to Gaussian random fields [16, 9].

If observation locations are regularly spaced on a rectangular grid with fixed spacing but increasing domain, it was shown by Davis and Borgman that the empirical semivariogram $\hat{\gamma}(h)$ is a consistent estimator for $\gamma(h)$, provided that the random field is m-dependent [10]. In this case, of course the estimator $\hat{\gamma}(h)$ is defined only when h is a possible distance between two points on the grid. The authors state this consistency result under the more general assumption that the semivariogram has finite range, mistakenly claiming that the m-dependence of the random field follows from this [10, p. 192]; for clarification on this point, see Example 1.1.

In the case where the observation locations are irregularly spaced, additional assumptions are required to ensure the consistency of the empirical semivariogram $\hat{\gamma}_\delta(h)$. In particular, an asymptotic framework is required in which not only the number of locations increases but the distance between them decreases, termed *mixed-increasing-domain asymptotics*. García et al. give some results on the asymptotic unbiasedness and consistency of the Nadaraya-Watson estimator $\hat{\gamma}_\delta^K(h)$ under such an asymptotic framework [14].

In light of the known results, the consistency of empirical semivariogram and related estimators is widely considered a settled matter. For example, Lahiri, Lee, and Cressie state [27],

> The simpler and more commonly used nonparametric estimators of the variogram, such as the method of moments estimator of Matheron (1962) and its robustified versions due to Cressie and Hawkins (1980) have many desirable properties like, unbiasedness, consistency, etc. ...

Regarding a kernel estimator of the covariance function, Hall and Patil remarked [18],

> It is not difficult to see that if, as n increases, the points t_i become increasingly dense in each bounded subset of $\mathbb{R}^d$, then the bandwidth h may be chosen so that $\check{\rho}(t) \to \rho(t)$ as $n \to \infty$, for each $t \in \mathbb{R}^d$.

However, in order to be true, such statements would need to be qualified by many assumptions on the random field as well as on the observation locations. We will see in §3.1 that even for well-behaved random fields (e.g., ρ^*-mixing Gaussian random fields),

it is not enough to assume that the observation locations become increasingly dense in each bounded subset; a stronger assumption must be made to ensure that the observation locations do not become denser in one region too much faster than in others. Applying the empirical semivariogram to data with heavily clustered observation locations may result in inconsistent estimation, a fact which appears to have gone unnoticed in the literature.

In Chapter 3, we will give consistency results for $\hat{\gamma}_\delta(h)$ under a more general version of the mixed-increasing-domain assumption than has previously been considered, simultaneously covering cases of both systematic and stochastic sampling designs, and we will establish not only pointwise covergence of $\hat{\gamma}_\delta(h)$ but uniform convergence for h on compact sets. In order to obtain such results, some form of assumption is needed that the dependence between two disjoint regions of a random field decays to zero as the distance between the regions increases. Assumptions along these lines, known as mixing conditions, are described in Chapter 2, along with new sufficient conditions which enable such mixing conditions to be established for certain types of Gaussian random fields. In Chapter 4, we give two applications of semivariogram estimation: an approach to digital image compression, and a refinement of Moran's I test for spatial correlation.

1.6 Proofs

Recall that a symmetric matrix $\Sigma = (\sigma_{ij}) \in \mathbb{R}^{n\times n}$ is said to be *positive semidefinite* if $\boldsymbol{a}^t\Sigma\boldsymbol{a} \geq 0$ for all vectors $\boldsymbol{a} \in \mathbb{R}^n$. Explicitly, this says that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij} \geq 0$$

for all $a_1, \ldots, a_n \in \mathbb{R}$. We will need a few elementary results about positive semidefinite matrices and multivariate Gaussian distributions:

Lemma 1.3. *Given a positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$, there is a unique positive semidefinite matrix $B \in \mathbb{R}^{n\times n}$ such that $A = B^2$.*

Proof. Since A is symmetric, by the spectral theorem there is an orthogonal matrix U and a real diagonal matrix D with $A = UDU^T$, and by replacing U with the product of U and suitable permutation matrix, we may assume without loss of generality that the diagonal entries of D are sorted in decreasing order. Let the diagonal entries of D be given by $d_1, \ldots, d_n$. Applying the positive semidefinite assumption on A with $\boldsymbol{a} = U\boldsymbol{e}_i$, we have

$$0 \leq \boldsymbol{a}^t A\boldsymbol{a} = \boldsymbol{e}_i^T U^T UDU^T U\boldsymbol{e}_i = \boldsymbol{e}_i^T D\boldsymbol{e}_i = d_i$$

Thus the diagonal entries of D are nonnegative, so we may define another real diagonal matrix D' with nonnegative diagonal entries given by $d'_i = \sqrt{d_i}$, and we have $(D')^2 = D$. If we set $B = UD'U^T$, it follows that

$$B^2 = UD'U^TUD'U^T = U(D')^2U^T = UDU^T = A,$$

which proves the existence part of the lemma. To prove uniqueness, suppose B_1 is another positive semidefinite matrix with $A = B_1^2$. Write $B_1 = U_1D_1U_1^T$, with the diagonal entries of D_1 sorted in decreasing order. By the argument above, the assumption that B_1 is positive semidefinite implies that the diagonal entries of D_1 are nonnegative. On the other hand, the diagonal entries of D_1 are the eigenvalues of B_1, and since $B_1^2 = A$ the squares of the eigenvalues of B_1 are the eigenvalues of A, which are the diagonal entries of D. It follows that D' and D_1 have the same diagonal entries, and since they are both sorted in decreasing order, we must have $D_1 = D'$. Now, the equation $B^2 = A = B_1^2$ implies that $DU^TU_1 = U^TU_1D$, i.e., the matrix U^TU_1 commutes with D. If the distinct diagonal entries of D are $\alpha_1, \ldots, \alpha_r$, with corresponding multiplicities $k_1, \ldots, k_r$, then we may write D as a direct sum of blocks:

$$D = \alpha_1 I_{k_1} \oplus \cdots \oplus \alpha_r I_{k_r}$$

The set of matrices which commute with D are those which are block diagonal, with block sizes $k_1, \ldots, k_r$, i.e., $\mathbb{R}^{k_1 \times k_1} \oplus \cdots \oplus R^{k_r \times k_r}$. On the other hand, D' may likewise be expressed

$$D' = \sqrt{\alpha_1} I_{k_1} \oplus \cdots \oplus \sqrt{\alpha_r} I_{k_r}$$

and since $\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_r}$ are distinct, the set of matrices commuting with D' is once again $\mathbb{R}^{k_1 \times k_1} \oplus \cdots \oplus R^{k_r \times k_r}$. Therefore, since U^TU_1 commutes with D, it also commutes with D', which means that $U^TU_1D' = D'U^TU_1$, hence $U_1D'U_1^T = UD'U^T$, i.e., $B_1 = B$. □

Lemma 1.4. *If $Z = (Z_1, \ldots, Z_n)$ is a vector of independent standard Gaussian random variables, and $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then QZ is also a vector of independent standard Gaussian random variables.*

Proof. The random vector Z has joint pdf $f(\boldsymbol{z}) = (2\pi)^{-n/2} e^{-\|\boldsymbol{z}\|^2/2}$. Since Q is orthogonal, $\|Q\boldsymbol{z}\| = \|\boldsymbol{z}\|$ for every $\boldsymbol{z} \in \mathbb{R}^n$, and since $|\det Q| = 1$, the change-of-variables formula shows that QZ has the same joint pdf as Z. Hence QZ is also a vector of independent standard Gaussian random variables. □

Theorem 1.5. *Given a symmetric matrix $\Sigma \in \mathbb{R}^{n \times n}$, there exists a random vector X in $\mathbb{R}^n$ with $\mathrm{Var}(X) = \Sigma$ if and only if Σ is positive semidefinite. Moreover, if Σ is positive*

semidefinite, then there is a unique multivariate Gaussian distribution with mean zero and covariance Σ*.*

Proof. Suppose $X = (X_1, \ldots, X_n)$ is a random vector with $\operatorname{Var}(X) = \Sigma$. Then for all $a_1, \ldots, a_n \in \mathbb{R}$,

$$0 \leq \operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{Cov}(X_i, X_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij}$$

so that Σ is positive semidefinite by definition. Conversely, suppose that Σ is positive semidefinite. Then there is a positive semidefinite matrix A with $\Sigma = A^2$. If we let $Z = (Z_1, \ldots, Z_n)$ be a vector of independent standard Gaussian random variables, and set $X = AZ$, then X is a multivariate Gaussian random vector with

$$\operatorname{Var}(X) = \operatorname{Var}(AZ) = A \operatorname{Var}(Z) A^T = AIA = A^2 = \Sigma$$

and mean zero, as desired.

To prove uniqueness, suppose we are given two matrices $A_1, A_2 \in \mathbb{R}^{n\times n}$, with $\operatorname{Var}(A_1 Z) = \Sigma = \operatorname{Var}(A_2 Z)$. Each matrix A_i, $i = 1, 2$, has a singular value decomposition $A_i = U_i D_i V_i^T$, where U_i and V_i are orthogonal matrices and D_i are diagonal matrices with nonnegative diagonal entries in descending order. Define $A_i' = U_i D_i U_i^T$. By Lemma 1.4, we have

$$A_i Z = U_i D_i V_i^T Z \overset{d}{=} U_i D_i Z \overset{d}{=} U_i D_i U_i^T Z = A_i' Z \tag{1.16}$$

We then have

$$\Sigma = \operatorname{Var}(A_i Z) = \operatorname{Var}(A_i' Z) = A_i'(A_i')^T = (A_i')^2$$

Since A_1' and A_2' are both then positive semidefinite square roots of Σ, Lemma 1.3 implies that $A_1' = A_2'$. From (1.16) it follows that

$$A_1 Z \overset{d}{=} A_1' Z = A_2' Z \overset{d}{=} A_2 Z,$$

as desired. □

We say that a function $C : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is *symmetric* if $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\boldsymbol{s}_2, \boldsymbol{s}_1)$ for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^n$.

Theorem 1.6. *Given a symmetric function $C : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, there exists a random field $Z(\boldsymbol{s})$ with covariance function C if and only if the inequality*

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j) \geq 0 \tag{1.17}$$

is satisfied for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^n$, and in this case Z may be taken to be a Gaussian random field with mean zero, and any two such Gaussian random fields have the same distribution.

Proof. Suppose there exists a random field $Z(\boldsymbol{s})$ with covariance function C. Then, for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^n$, we have

$$0 \leq \operatorname{Var}\left(\sum_{i=1}^{k} a_i Z(\boldsymbol{s}_i) \right) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \operatorname{Cov}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j)$$

so the inequality (1.17) holds. Conversely, let C be given and assume that (1.17) holds. By Theorem 1.5, this means that for every finite subset $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\} \subseteq \mathbb{R}^n$, there is a mean zero Gaussian random vector on $\mathbb{R}^k$ with covariance matrix $(C(\boldsymbol{s}_i, \boldsymbol{s}_j))$. Since a multivariate Gaussian distribution is uniquely determined by its mean and covariance, we may apply the Kolmogorov extension theorem [22, Theorem 6.16] to conclude that there exists a Gaussian random field $Z(\boldsymbol{s})$ such that the restriction of Z to any finite subset $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k\} \subseteq \mathbb{R}^n$ has a multivariate Gaussian distribution with mean zero and covariance $(C(\boldsymbol{s}_i, \boldsymbol{s}_j))$. In particular, by restricting to the subset $\{\boldsymbol{s}_1, \boldsymbol{s}_2\}$, we see that $\operatorname{Cov}(Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)) = C(\boldsymbol{s}_1, \boldsymbol{s}_2)$, so that Z has covariance function C, as required. The last statement is a consequence of the facts that the joint distribution of a random field is determined by its finite-dimensional distributions [22, Proposition 3.2], and that by Theorem 1.5 a (finite-dimensional) multivariate Gaussian distribution with mean zero is uniquely determined by its covariance matrix. □

We say that a function $C : \mathbb{R}^n \to \mathbb{R}$ is *positive semidefinite* if

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i - \boldsymbol{s}_j) \geq 0$$

for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k \in \mathbb{R}^n$. Cressie [9, §2.5.1] sketches a proof that this condition characterizes the even functions which are covariance functions of second-order stationary random fields. However, Cressie's proof relies on a spectral representation of $Z(\boldsymbol{s})$, which exists only if $C(\boldsymbol{h})$ is continuous. Nevertheless, this characterization is correct even if $C(\boldsymbol{h})$ is not necessarily continuous:

Theorem 1.7. *Given a function $C : \mathbb{R}^n \to \mathbb{R}$, there exists some second-order stationary random field $Z(\boldsymbol{s})$ on $\mathbb{R}^n$ with covariance function C if and only if C is a positive semidefinite even function.*

Proof. Since a random field has stationary covariance function $C(\boldsymbol{h})$ if and only if it has covariance function $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = C(\boldsymbol{s}_1 - \boldsymbol{s}_2)$, and since the function $C(\boldsymbol{s}_1, \boldsymbol{s}_2)$ is symmetric if and only if $C(\boldsymbol{h})$ is even, the result follows immediately from Theorem 1.6. □

A complex-valued function $C : \mathbb{R}^n \to \mathbb{C}$ is said to be *positive semidefinite* if

$$\sum_{j=1}^{m}\sum_{k=1}^{m} a_j\overline{a_k}C(\boldsymbol{s}_j - \boldsymbol{s}_k) \geq 0$$

for every $m \in \mathbb{N}$, $a_1, \ldots, a_m \in \mathbb{C}$, and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m \in \mathbb{R}^n$. By composition with the inclusion $\mathbb{R} \to \mathbb{C}$, any real-valued function $C(\boldsymbol{h})$ may also be considered as a complex-valued function, in which case we have two competing definitions of what it means for C to be positive semidefinite. However, if $C(\boldsymbol{h})$ is an even function, these two definitions coincide:

Theorem 1.8. *An even function $C : \mathbb{R}^n \to \mathbb{R}$ is positive semidefinite (as a real-valued function) if an only if it is positive semidefinite as a complex-valued function.*

Proof. The "if" part of the statement is trivial, so assume that $C(\boldsymbol{h})$ is positive semidefinite as a real-valued function. Given $m \in \mathbb{N}$, $a_1, \ldots, a_m \in \mathbb{C}$, and $s_1, \ldots, s_m \in \mathbb{R}^n$, we may write $a_j = b_j + c_j i$ with $b_j, c_j \in \mathbb{R}$. Then, writing $C_{jk} = C(\boldsymbol{s}_j - \boldsymbol{s}_k)$, since C is even we have $C_{jk} = C_{kj}$ and hence

$$\begin{aligned}
&\sum_{j=1}^{m}\sum_{k=1}^{m} a_j\overline{a_k}C(\boldsymbol{s}_j - \boldsymbol{s}_k) \\
&= \sum_{j=1}^{m}\sum_{k=1}^{m}(b_j + c_j i)(b_k - c_k i)C_{jk} \\
&= \sum_{j=1}^{m}\sum_{k=1}^{m} b_j b_k C_{jk} + \sum_{j=1}^{m}\sum_{k=1}^{m} c_j c_k C_{jk} - i\sum_{j=1}^{m}\sum_{k=1}^{m} b_j c_k C_{jk} + i\sum_{j=1}^{m}\sum_{k=1}^{m} b_k c_j C_{jk} \\
&= \sum_{j=1}^{m}\sum_{k=1}^{m} b_j b_k C_{jk} + \sum_{j=1}^{m}\sum_{k=1}^{m} c_j c_k C_{jk} \\
&\geq 0
\end{aligned}$$

Thus C is positive semidefinite as a complex-valued function, as required. □

We say that a function $\gamma : \mathbb{R}^n \to \mathbb{R}$ is *conditionally negative semidefinite* if

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j) \leq 0$$

for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$, such that $\sum_{i=1}^{k} a_i = 0$. We can give a similar definition for complex-valued functions: we say that a function $\gamma : \mathbb{R}^n \to \mathbb{C}$ is *conditionally negative semidefinite* if

$$\sum_{i=1}^{k} \sum_{j=1}^{k} a_i \overline{a_j} \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j) \leq 0$$

for every $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{C}$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$, such that $\sum_{i=1}^{k} a_i = 0$. The following statement is then analogous to Theorem 1.8 and is proven in exactly the same way:

Theorem 1.9. *An even function $\gamma : \mathbb{R}^n \to \mathbb{R}$ is conditionally negative semidefinite (as a real-valued function) if an only if it is conditionally negative semidefinite as a complex-valued function.*

We have the following characterization of which functions can occur as the semivariogram of some intrinsically stationary random field:

Theorem 1.10. *Given a function $\gamma : \mathbb{R}^n \to \mathbb{R}$, there exists an intrinsically stationary random field $Z(\boldsymbol{s})$ on $\mathbb{R}^n$ with semivariogram γ if and only if γ is a conditionally negative semidefinite even function with $\gamma(\mathbf{0}) = 0$. Moreover, given such a function γ,*

(a) There is a unique (in distribution) Gaussian random field $Z_0(\boldsymbol{s})$ with mean zero and semivariogram γ such that $Z_0(\boldsymbol{s}) = 0$ almost surely. The covariance function of such a random field is given by

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \gamma(\boldsymbol{s}_1) + \gamma(\boldsymbol{s}_2) - \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_2)$$

(b) Any Gaussian random field $Z(\boldsymbol{s})$ with mean zero and semivariogram γ can be decomposed as

$$Z(\boldsymbol{s}) = Z_0(\boldsymbol{s}) + Y$$

for some random field Z_0 with $Z_0(\mathbf{0}) = 0$ almost surely (hence having distribution uniquely determined by (a)) and some random variable Y (not necessarily independent of Z_0).

Proof. Suppose that γ is the semivariogram of a random field $Z(\boldsymbol{s})$. First note that for any $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, we have

$$\begin{aligned}\gamma(\boldsymbol{s}_1-\boldsymbol{s}_2) &= \frac{1}{2}\operatorname{Var}(Z(\boldsymbol{s}_1)-Z(\boldsymbol{s}_2))\\ &= \frac{1}{2}\operatorname{Var}\big((Z(\boldsymbol{s}_1)-Z(\mathbf{0}))-(Z(\boldsymbol{s}_2)-Z(\mathbf{0}))\big)\\ &= \frac{1}{2}\left[\operatorname{Var}(Z(\boldsymbol{s}_1)-Z(\mathbf{0}))+\operatorname{Var}(\boldsymbol{s}_2)-2\operatorname{Cov}(Z(\boldsymbol{s}_2)-Z(\mathbf{0}),Z(\boldsymbol{s}_1)-Z(\mathbf{0}))\right]\\ &= \gamma(\boldsymbol{s}_1)+\gamma(\boldsymbol{s}_2)-\operatorname{Cov}(Z(\boldsymbol{s}_1)-Z(\mathbf{0}),Z(\boldsymbol{s}_2)-Z(\mathbf{0}))\end{aligned}$$

so that

$$\operatorname{Cov}(Z(\boldsymbol{s}_1)-Z(\mathbf{0}),Z(\boldsymbol{s}_2)-Z(\mathbf{0}))=\gamma(\boldsymbol{s}_1)+\gamma(\boldsymbol{s}_2)-\gamma(\boldsymbol{s}_1-\boldsymbol{s}_2) \tag{1.18}$$

Now, given $k\in\mathbb{N}$, $a_1,\dots,a_k\in\mathbb{R}$, and $\boldsymbol{s}_1,\dots,\boldsymbol{s}_k\in\mathbb{R}^n$ such that $\sum_{i=1}^k a_i=0$, we have

$$\begin{aligned}0 &\le \operatorname{Var}\left(\sum_{i=1}^k a_i(Z(\boldsymbol{s}_i)-Z(\mathbf{0}))\right)\\ &= \sum_{i=1}^k\sum_{j=1}^k a_ia_j\operatorname{Cov}(Z(\boldsymbol{s}_i)-Z(\mathbf{0}),Z(\boldsymbol{s}_j)-Z(\mathbf{0}))\\ &= \sum_{i=1}^k\sum_{j=1}^k a_ia_j(\gamma(\boldsymbol{s}_i)+\gamma(\boldsymbol{s}_j)-\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j))\\ &= \left(\sum_{j=1}^k a_j\right)\sum_{i=1}^k a_i\gamma(\boldsymbol{s}_i)+\left(\sum_{i=1}^k a_i\right)\sum_{j=1}^k a_j\gamma(\boldsymbol{s}_j)-\sum_{i=1}^k\sum_{j=1}^k a_ia_j\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j)\\ &= -\sum_{i=1}^k\sum_{j=1}^k a_ia_j\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j)\end{aligned}$$

which shows that γ is conditionally negative definite. It is trivial to check that γ is an even function and that $\gamma(\mathbf{0})=0$.

Now suppose that $Z_0(\boldsymbol{s})$ is a Gaussian random field with mean zero and semivariogram γ such that $Z_0(\mathbf{0})=0$ almost surely. Then (1.18) becomes

$$C(\boldsymbol{s}_1,\boldsymbol{s}_2)=\gamma(\boldsymbol{s}_1)+\gamma(\boldsymbol{s}_2)-\gamma(\boldsymbol{s}_1-\boldsymbol{s}_2) \tag{1.19}$$

and the distribution of Z_0 is determined by Theorem 1.6, which proves the uniqueness in (a). To prove the existence in (a), define $C(\boldsymbol{s}_1,\boldsymbol{s}_2)$ by (1.19), and let $k\in\mathbb{N}$, $a_1,\dots,a_k\in\mathbb{R}$, and $\boldsymbol{s}_1,\dots,\boldsymbol{s}_k\in\mathbb{R}^n$ be given. Since γ is even, C is symmetric. Setting $a_0=-\sum_{i=1}^k a_i$ and $\boldsymbol{s}_0=\mathbf{0}$, the assumptions on γ together with (1.19) imply that

$$0\ge\sum_{i=0}^k\sum_{j=0}^k a_ia_j\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j)$$

$$
\begin{aligned}
&= a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) + a_0 \sum_{j=1}^{k} a_j \gamma(-\boldsymbol{s}_j) + \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \\
&= 2a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) + \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) \\
&= 2a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) + \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j (\gamma(\boldsymbol{s}_i) + \gamma(\boldsymbol{s}_j) - C(\boldsymbol{s}_i, \boldsymbol{s}_j)) \\
&= 2a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) + \left(\sum_{j=1}^{k} a_j \right) \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) + \left(\sum_{i=1}^{k} a_i \right) \sum_{i=j}^{k} a_j \gamma(\boldsymbol{s}_j) - \sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j) \\
&= 2a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) - a_0 \sum_{i=1}^{k} a_i \gamma(\boldsymbol{s}_i) - a_0 \sum_{i=j}^{k} a_j \gamma(\boldsymbol{s}_j) - \sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j)) \\
&= -\sum_{j=1}^{k} a_i a_j C(\boldsymbol{s}_i, \boldsymbol{s}_j)
\end{aligned}
$$

so that C is positive semidefinite. Theorem 1.6 implies that there exists a random field $Z_0(\boldsymbol{s})$ with mean zero and covariance function $C(\boldsymbol{s}_1, \boldsymbol{s}_2)$. For such a random field we have

$$\mathrm{Var}(Z(\mathbf{0})) = C(\mathbf{0}, \mathbf{0}) = \gamma(\mathbf{0}) + \gamma(\mathbf{0}) - \gamma(\mathbf{0} - \mathbf{0}) = 0,$$

so that $Z(\mathbf{0}) = 0$ almost surely. This completes the proof of the main statement of the theorem and (a).

Given a Gaussian random field $Z(\boldsymbol{s})$ with mean zero and semivariogram γ, if we define Z_0 by $Z_0(\boldsymbol{s}) = Z(\boldsymbol{s}) - Z(\mathbf{0})$, then $Z_0(\mathbf{0}) = 0$, so (b) follows immediately, by setting $Y = Z(\mathbf{0})$. □

A continuous function $g : [0, \infty) \to \mathbb{R}$ is *completely monotone* if g is infinitely differentiable on $(0, \infty)$ and

$$(-1)^n \frac{d^n g(t)}{dt^n} \geq 0$$

for all $t > 0$ and all $n \in \mathbb{N}$.

We will use the following result from Schoenberg [37, Theorem 3]:

Theorem 1.11. *A function $g : [0, \infty) \to \mathbb{R}$ is completely monotone if and only if the function $C(\boldsymbol{h}) = g(\|\boldsymbol{h}\|^2)$ for $\boldsymbol{h} \in \mathbb{R}^n$ is positive semidefinite for all $n = 1, 2, \ldots$.*

Theorem 1.12. *The functions $e^{-\|\boldsymbol{h}\|}$ and $e^{-\|\boldsymbol{h}\|^2}$ on $\boldsymbol{h} \in \mathbb{R}^n$ are positive semidefinite.*

Proof. First consider $C(\boldsymbol{h}) = e^{-\|\boldsymbol{h}\|^2}$. Applying Theorem 1.11 with $g(t) = e^{-t}$, it suffices to show that $g(t)$ is completely monotone, but this is clear since

$$(-1)^n \frac{d^n g(t)}{dt^n} = e^{-t} \geq 0.$$

Now consider $C(\boldsymbol{h}) = e^{-\|\boldsymbol{h}\|}$. Applying Theorem 1.11 with $g(t) = e^{-t^{1/2}}$, it suffices to show that $g(t)$ is completely monotone. We claim that $\frac{d^n g(t)}{dt^n}$ has the form

$$\frac{d^n g(t)}{dt^n} = (-1)^n \sum_{i=1}^{k_n} c_{n,i} t^{-\alpha_{n,i}} e^{-t^{1/2}}$$

for positive constants $c_{n,i}$ and $\alpha_{n,i}$. This is clearly true for $n = 0$, and assuming by induction that the statement is true for a certain n, we then have

$$\begin{aligned}
\frac{d^{n+1} g(t)}{dt^{n+1}} &= \frac{d}{dt}\left(\frac{d^n g(t)}{dt^n}\right) \\
&= \frac{d}{dt}\left((-1)^n \sum_{i=1}^{k_n} c_{n,i} t^{-\alpha_{n,i}} e^{-t^{1/2}}\right) \\
&= (-1)^n \sum_{i=1}^{k_n} c_{n,i} \frac{d}{dt}\left(t^{-\alpha_{n,i}} e^{-t^{1/2}}\right) \\
&= (-1)^n \sum_{i=1}^{k_n} c_{n,i} \left(-\alpha_{n,i} t^{-\alpha_{n,i}-1} e^{-t^{1/2}} - \frac{1}{2} t^{-\alpha_{n,i}-\frac{1}{2}} e^{-t^{1/2}}\right)
\end{aligned}$$

which again has the desired form, proving the claim, and it follows immediately that $g(t)$ is completely monotone, as required. □

Combining Theorems 1.12 and 1.2, we deduce the following:

Theorem 1.13. *The function $\gamma(\boldsymbol{h}) = \|\boldsymbol{h}\|$ is conditionally negative semidefinite.*

Lemma 1.14. *Let $g(\boldsymbol{s})$ be real-valued locally integrable function defined on $\mathbb{R}^n$. Assume that for every bounded Borel measurable set $A \subset \mathbb{R}^n$,*

$$\int_A g(\boldsymbol{s})\ d\boldsymbol{s} = 0$$

Then $g(\boldsymbol{s}) = 0$ almost everywhere.

Proof. Consider the open ball B_r of radius r centered at the origin. Define

$$A_r^+ = \{\boldsymbol{s} \in B_r : g(\boldsymbol{s}) \geq 0\}$$
$$A_r^- = \{\boldsymbol{s} \in B_r : g(\boldsymbol{s}) < 0\}$$

Since A_r^+ is a bounded Borel measurable set, by assumption we have $\int_{A_r^+} g(\boldsymbol{s})\, d\boldsymbol{s} = 0$. Since $g(\boldsymbol{s})$ is nonnegative on A_r^+, this implies that $g(\boldsymbol{s}) = 0$ almost everywhere on A_r^+. Similarly $g(\boldsymbol{s}) = 0$ almost everywhere on A_r^-, so that $g(\boldsymbol{s}) = 0$ almost everywhere on $B_r = A_r^+ \cup A_r^-$. Since $\mathbb{R}^n$ can be expressed as the countable union $\mathbb{R}^n = \cup_{r=1}^{\infty} B_r$, it follows that $g(\boldsymbol{s}) = 0$ almost everywhere on $\mathbb{R}^n$. □

The following result may be deduced from [34, Theorem 9.5], but for completeness we include a proof:

Lemma 1.15. *Let $g(\boldsymbol{s})$ be a real-valued locally integrable function defined on $\mathbb{R}^n$. For any bounded Borel measurable set $A \subseteq \mathbb{R}^n$, the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by*

$$f(\boldsymbol{h}) = \int_A |g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s})|\, d\boldsymbol{s}$$

is continuous.

Proof. Let $\boldsymbol{h}_0 \in \mathbb{R}^n$ and $\epsilon > 0$ be given. We have

$$\begin{aligned}
|f(\boldsymbol{h}) - f(\boldsymbol{h}_0)| &= \left| \int_A |g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s})|\, d\boldsymbol{s} - \int_A |g(\boldsymbol{s}+\boldsymbol{h}_0) - g(\boldsymbol{s})|\, d\boldsymbol{s} \right| \\
&= \left| \int_A (|g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s})| - |g(\boldsymbol{s}+\boldsymbol{h}_0) - g(\boldsymbol{s})|)\, d\boldsymbol{s} \right| \\
&\leq \int_A \Big| |g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s})| - |g(\boldsymbol{s}+\boldsymbol{h}_0) - g(\boldsymbol{s})| \Big|\, d\boldsymbol{s} \\
&\leq \int_A |g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s}+\boldsymbol{h}_0)|\, d\boldsymbol{s}.
\end{aligned}$$

Let U be the open neighborhood of $A + \boldsymbol{h}_0$ defined by $U = A + \boldsymbol{h}_0 + B_1$, where B_1 is the standard unit ball. Since $C_c(U)$ is dense in $L^1(U)$, we may choose a function $g_0 \in C_c(U)$ such that $\int_U |g(\boldsymbol{s}) - g_0(\boldsymbol{s})|\, d\boldsymbol{s} < \epsilon/3$. Since g_0 is uniformly continuous, there exists δ with $0 < \delta < 1$ such that $|g_0(\boldsymbol{s}_2) - g_0(\boldsymbol{s}_1)| < \frac{\epsilon}{3}(\lambda(U))^{-1}$ for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in U$. In particular, this implies $|g_0(\boldsymbol{s}+\boldsymbol{h}) - g_0(\boldsymbol{s})| \leq \frac{\epsilon}{3}(\lambda(U))^{-1}$ for all $\boldsymbol{s} \in A$ and all $\boldsymbol{h}$ such that $|\boldsymbol{h} - \boldsymbol{h}_0| < \delta$. Then for $\|\boldsymbol{h} - \boldsymbol{h}_0\| < \delta$,

$$\begin{aligned}
|f(\boldsymbol{h}) - f(\boldsymbol{h}_0)| &\leq \int_A |g(\boldsymbol{s}+\boldsymbol{h}) - g(\boldsymbol{s}+\boldsymbol{h}_0)|\, d\boldsymbol{s} \\
&\leq \int_A |g(\boldsymbol{s}+\boldsymbol{h}) - g_0(\boldsymbol{s}+\boldsymbol{h})|\, d\boldsymbol{s} + \int_A |g_0(\boldsymbol{s}+\boldsymbol{h}) - g_0(\boldsymbol{s}+\boldsymbol{h}_0)|\, d\boldsymbol{s} \\
&\quad + \int_A |g_0(\boldsymbol{s}+\boldsymbol{h}_0) - g(\boldsymbol{s}+\boldsymbol{h}_0)|\, d\boldsymbol{s} \\
&\leq \int_U |g(\boldsymbol{s}) - g_0(\boldsymbol{s})|\, d\boldsymbol{s} + \int_A \frac{\epsilon}{3}(\lambda(U))^{-1}\, d\boldsymbol{s} + \int_U |g_0(\boldsymbol{s}) - g(\boldsymbol{s})|\, d\boldsymbol{s} \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}
\end{aligned}$$

$$= \epsilon$$

This proves that $f(\boldsymbol{h})$ is continuous. □

In order to prove Theorem 1.17, we need a technical lemma characterizing the continuous functions which are the absolute value of a linear functional (related to a result of Jarczyk and Volkmann [20]):

Lemma 1.16. *Let $q(\boldsymbol{h})$ be a nonnegative, continuous real-valued function defined on $\mathbb{R}^n$. Assume that $q(\boldsymbol{h})$ satisfies the condition*

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2) \in \{q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2), |q(\boldsymbol{h}_1) - q(\boldsymbol{h}_2)|\} \tag{1.20}$$

for all $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^n$. Then there exists a linear functional $L : \mathbb{R}^n \to \mathbb{R}$ such that $q(\boldsymbol{h}) = |L(\boldsymbol{h})|$ for all $\boldsymbol{h} \in \mathbb{R}^n$.

Proof. If q is identically zero, then the conclusion follows trivially by taking $L(\boldsymbol{h}) = 0$. So without loss of generality we may choose some $\boldsymbol{h}_0 \in \mathbb{R}^n$ with $q(\boldsymbol{h}_0) > 0$. Define

$$\begin{aligned} H_+ &= \{\boldsymbol{h} \in \mathbb{R}^n : q(\boldsymbol{h} + \boldsymbol{h}_0) = q(\boldsymbol{h}) + q(\boldsymbol{h}_0)\} \\ H_- &= \{\boldsymbol{h} \in \mathbb{R}^n : q(\boldsymbol{h} + \boldsymbol{h}_0) = |q(\boldsymbol{h}) - q(\boldsymbol{h}_0)|\} \\ H_0 &= \{\boldsymbol{h} \in \mathbb{R}^n : q(\boldsymbol{h}) = 0\} \end{aligned}$$

The assumption (1.20) implies that $\mathbb{R}^n = H_+ \cup H_-$. We claim that $H_+ \cap H_- = H_0$. If $\boldsymbol{h} \in H_+ \cap H_-$, then either $q(\boldsymbol{h}) + q(\boldsymbol{h}_0) = q(\boldsymbol{h}) - q(\boldsymbol{h}_0)$ or $q(\boldsymbol{h}) + q(\boldsymbol{h}_0) = q(\boldsymbol{h}_0) - q(\boldsymbol{h})$, but the first case is impossible since $q(\boldsymbol{h}_0) \neq 0$, so it follows that $q(\boldsymbol{h}) = 0$, i.e., $\boldsymbol{h} \in H_0$. Conversely, if $\boldsymbol{h} \in H_0$ then $q(\boldsymbol{h}) + q(\boldsymbol{h}_0) = q(\boldsymbol{h}_0) = |q(\boldsymbol{h}) - q(\boldsymbol{h}_0)|$, so $\boldsymbol{h} \in H_+ \cap H_-$, proving the claim.

Note that since q is continuous, H_+, H_-, and H_0 are closed subsets of $\mathbb{R}^n$. Therefore, if we define

$$H'_+ = H_+ \setminus H_0, \quad H'_- = H_- \setminus H_0$$

then H'_+ and H'_- are disjoint and closed in $\mathbb{R}^n \setminus H_0$, so that $\mathbb{R}^n \setminus H_0 = H'_+ \cup H'_-$ is a disconnection of $\mathbb{R}^n \setminus H_0$.

Now define

$$L(\boldsymbol{h}) = \begin{cases} q(\boldsymbol{h}), & \text{if } \boldsymbol{h} \in H_+ \\ -q(\boldsymbol{h}), & \text{if } \boldsymbol{h} \in H_- \end{cases}$$

This is well-defined since on $H_+ \cap H_- = H_0$ we have $q(\boldsymbol{h}) = -q(\boldsymbol{h}) = 0$. Clearly $q(\boldsymbol{h}) = |L(\boldsymbol{h})|$. It only remains to prove that $L(\boldsymbol{h})$ is linear. We proceed in steps by proving several claims:

Claim 1. $q(\mathbf{0}) = 0$.

By applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h}_2 = \mathbf{0}$, we deduce that $q(\mathbf{0}) \in \{2q(\mathbf{0}), 0\}$. In both cases, we conclude $q(\mathbf{0}) = 0$.

Claim 2. For all $\boldsymbol{h} \in \mathbb{R}^n$,

$$q(-\boldsymbol{h}) = q(\boldsymbol{h}). \tag{1.21}$$

Applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h}$ and $\boldsymbol{h}_2 = -\boldsymbol{h}$, we deduce that $0 = q(\mathbf{0}) \in \{q(\boldsymbol{h}) + q(-\boldsymbol{h}), |q(\boldsymbol{h}) - q(-\boldsymbol{h})\}$. If $0 = q(\boldsymbol{h}) + q(-\boldsymbol{h})$, then the nonnegativity of q implies that $q(\boldsymbol{h}) = 0 = q(-\boldsymbol{h})$. Therefore, in both cases we have $q(-\boldsymbol{h}) = q(\boldsymbol{h})$.

Claim 3. For all real numbers $\alpha \geq 0$ and all $\boldsymbol{h} \in \mathbb{R}^n$,

$$q(\alpha \boldsymbol{h}) = \alpha q(\boldsymbol{h}). \tag{1.22}$$

The claim is clearly true for $\alpha = 0, 1$. If the claim holds for a particular $\boldsymbol{h}$ with $\alpha = 2$ and $\alpha = 3$, we will first use induction to show that it holds for this $\boldsymbol{h}$ for all $\alpha \in \mathbb{N}$. Assume without loss of generality that $q(\boldsymbol{h}) \neq 0$ (otherwise the claim is trivial), fix a natural number $k \geq 3$, and suppose that the claim holds for a particular $\boldsymbol{h}$ for all natural numbers $\alpha \leq k$. Applying (1.20) with $\boldsymbol{h}_1 = k\boldsymbol{h}$ and $\boldsymbol{h}_2 = \boldsymbol{h}$, and using the fact that $q(k\boldsymbol{h}) = kq(\boldsymbol{h})$, we have that

$$q((k+1)\boldsymbol{h}) \in \{(k+1)q(\boldsymbol{h}), (k-1)q(\boldsymbol{h}), (1-k)q(\boldsymbol{h})\}. \tag{1.23}$$

However, the possibility $q((k+1)\boldsymbol{h}) = (1-k)q(\boldsymbol{h})$ is excluded since $k > 1$ and q is nonnegative. Applying (1.20) with $\boldsymbol{h}_1 = (k-1)\boldsymbol{h}$ and $\boldsymbol{h}_2 = 2\boldsymbol{h}$, and using the fact that $q((k-1)\boldsymbol{h}) = (k-1)q(\boldsymbol{h})$ and $q(2\boldsymbol{h}) = 2q(\boldsymbol{h})$, we have that

$$q((k+1)\boldsymbol{h}) \in \{(k+1)\boldsymbol{h}, (k-3)q(\boldsymbol{h}), (3-k)q(\boldsymbol{h})\}. \tag{1.24}$$

Comparing (1.23) and (1.24), if $q((k+1)\boldsymbol{h}) \neq (k+1)q(\boldsymbol{h})$, then since $q(\boldsymbol{h}) \neq 0$ we must have $k - 1 = 3 - k$, which is possible only if $k = 2$, contrary to assumption. Therefore, we must have $q((k+1)\boldsymbol{h}) = (k+1)q(\boldsymbol{h})$. This shows that for any particular $\boldsymbol{h}$, the claim will hold for all $\alpha \in \mathbb{N}$ if it holds for $\alpha = 2$ and $\alpha = 3$.

Now let $\boldsymbol{h}$ be given with $q(\boldsymbol{h}) \neq 0$. Applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h}_2 = \frac{1}{2}\boldsymbol{h}$, we deduce that $q(\boldsymbol{h}) \in \{2q\left(\frac{1}{2}\boldsymbol{h}\right), 0\}$. But since $q(\boldsymbol{h}) \neq 0$, we must have $q\left(\frac{1}{2}\boldsymbol{h}\right) = \frac{1}{2}q(\boldsymbol{h})$. Applying this same argument again with $\frac{1}{2}\boldsymbol{h}$ in place of $\boldsymbol{h}$, we must have $q\left(\frac{1}{4}\boldsymbol{h}\right) = \frac{1}{4}q(\boldsymbol{h})$. Now, applying (1.20) with $\boldsymbol{h}_1 = \frac{1}{2}\boldsymbol{h}$ and $\boldsymbol{h}_2 = \frac{1}{4}\boldsymbol{h}$, we deduce that $q\left(\frac{3}{4}\boldsymbol{h}\right) \in \left\{\frac{3}{4}q(\boldsymbol{h}), \frac{1}{4}q(\boldsymbol{h})\right\}$. But if $q\left(\frac{3}{4}\boldsymbol{h}\right) = \frac{1}{4}q(\boldsymbol{h})$, then applying (1.20) with $\boldsymbol{h}_1 = \frac{3}{4}\boldsymbol{h}$ and $\boldsymbol{h}_2 = \frac{1}{4}\boldsymbol{h}$ gives $q(\boldsymbol{h}) \in \left\{\frac{1}{2}q(\boldsymbol{h}), 0\right\}$, contradicting that $q(\boldsymbol{h}) \neq 0$. So we must have $q\left(\frac{3}{4}\boldsymbol{h}\right) = \frac{3}{4}q(\boldsymbol{h})$. Now we have $q\left(\alpha\left(\frac{1}{4}\boldsymbol{h}\right)\right) =$

$\alpha q\left(\frac{1}{4}\boldsymbol{h}\right)$ for $\alpha = 2$ and $\alpha = 3$, so that by the preceding argument $q\left(\alpha\left(\frac{1}{4}\boldsymbol{h}\right)\right) = \alpha q\left(\frac{1}{4}\boldsymbol{h}\right)$ for all natural numbers α. It follows that

$$q(\alpha\boldsymbol{h}) = q\left(4\alpha\left(\frac{1}{4}\boldsymbol{h}\right)\right) = 4\alpha q\left(\frac{1}{4}\boldsymbol{h}\right) = \alpha q(\boldsymbol{h})$$

for all natural numbers α. Substituting $\frac{1}{\alpha}\boldsymbol{h}$ in place of $\boldsymbol{h}$, we obtain $q\left(\frac{1}{\alpha}\boldsymbol{h}\right) = \frac{1}{\alpha}q(\boldsymbol{h})$ for natural numbers $\alpha > 0$. It follows that for any rational number $\alpha = \frac{a}{b}$,

$$q(\alpha\boldsymbol{h}) = q\left(a\left(\frac{1}{b}\boldsymbol{h}\right)\right) = aq\left(\frac{1}{b}\boldsymbol{h}\right) = \frac{a}{b}q(\boldsymbol{h}) = \alpha q(\boldsymbol{h})$$

Now since $q(\alpha\boldsymbol{h}) - \alpha q(\boldsymbol{h})$ is a continuous function of α which is zero for all $\alpha \in \mathbb{Q}$, it must be zero for all $\alpha \in \mathbb{R}$, proving the claim.

Claim 4. The equality $-H_+ = H_-$ holds.

To prove this, let $\boldsymbol{h} \in H_+$ be given. Then either $-\boldsymbol{h} \in H_+$ or $-\boldsymbol{h} \in H_-$. Suppose $-\boldsymbol{h} \in H_+$. Then by the definition of H_+ and (1.21) we have

$$q(\boldsymbol{h} + \boldsymbol{h}_0) = q(\boldsymbol{h}) + q(\boldsymbol{h}_0)$$
$$q(-\boldsymbol{h} + \boldsymbol{h}_0) = q(-\boldsymbol{h}) + q(\boldsymbol{h}_0) = q(\boldsymbol{h}) + q(\boldsymbol{h}_0)$$

Applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h} + \boldsymbol{h}_0$ and $\boldsymbol{h}_2 = \boldsymbol{h} - \boldsymbol{h}_0$ we deduce that either $q(2\boldsymbol{h}_0) = 2q(\boldsymbol{h}) + 2q(\boldsymbol{h}_0)$ or $q(2\boldsymbol{h}_0) = 0$. The latter case is impossible since $q(2\boldsymbol{h}_0) = 2q(\boldsymbol{h}_0) \neq 0$ by the choice of $\boldsymbol{h}_0$. In the former case, we have $q(\boldsymbol{h}) = 0$ and hence $q(-\boldsymbol{h}) = -q(\boldsymbol{h}) = 0$, i.e., $-\boldsymbol{h} \in H_0 \subseteq H_-$. This proves that $-H_+ \subseteq H_-$.

Now let $\boldsymbol{h} \in H_-$ be given, and suppose that $-\boldsymbol{h} \in H_-$. Then by the definition of H_- and (1.21) we have

$$q(\boldsymbol{h} + \boldsymbol{h}_0) = |q(\boldsymbol{h}) - q(\boldsymbol{h}_0)| \tag{1.25}$$
$$q(-\boldsymbol{h} + \boldsymbol{h}_0) = |q(-\boldsymbol{h}) - q(\boldsymbol{h}_0)| = |q(\boldsymbol{h}) - q(\boldsymbol{h}_0)| \tag{1.26}$$

Applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h} + \boldsymbol{h}_0$ and $\boldsymbol{h}_2 = \boldsymbol{h} - \boldsymbol{h}_0$ we deduce that either $q(2\boldsymbol{h}_0) = 2|q(\boldsymbol{h}) - q(\boldsymbol{h}_0)|$ or $q(2\boldsymbol{h}_0) = 0$. The latter case is impossible, and the former case splits into two subcases: either $q(2\boldsymbol{h}_0) = 2q(\boldsymbol{h}) - 2q(\boldsymbol{h}_0)$ or $q(2\boldsymbol{h}_0) = 2q(\boldsymbol{h}_0) - 2q(\boldsymbol{h})$. In the latter case, $q(\boldsymbol{h}) = 0$ so that also $q(-\boldsymbol{h}) = 0$ and hence $-\boldsymbol{h} \in H_0 \subseteq H_+$. So assume we are in the remaining case, $q(2\boldsymbol{h}_0) = 2q(\boldsymbol{h}) - 2q(\boldsymbol{h}_0)$, in which case $q(\boldsymbol{h}) = 2q(\boldsymbol{h}_0)$. Applying (1.21) to (1.26), we have

$$q(\boldsymbol{h} - \boldsymbol{h}_0) = |q(\boldsymbol{h}) - q(\boldsymbol{h}_0)| \tag{1.27}$$

Applying (1.20) with $\boldsymbol{h}_1 = \boldsymbol{h} + \boldsymbol{h}_0$ and $\boldsymbol{h}_2 = \boldsymbol{h} - \boldsymbol{h}_0$ and substituting (1.25) and (1.27), we deduce that either $q(2\boldsymbol{h}) = 2|q(\boldsymbol{h}) - q(\boldsymbol{h}_0)|$ or $q(2\boldsymbol{h}) = 0$. In the latter case, $\boldsymbol{h} \in H_0 \subseteq -H_+$,

so assume the former case, which splits into two subcases: either $q(2\boldsymbol{h}) = 2q(\boldsymbol{h}) - 2q(\boldsymbol{h}_0)$ or $q(2\boldsymbol{h}) = 2q(\boldsymbol{h}_0) - 2q(\boldsymbol{h})$. The former case is impossible since $q(\boldsymbol{h}_0) \neq 0$, so we must have the latter case, and $q(\boldsymbol{h}) = \frac{1}{2}q(\boldsymbol{h}_0)$, but this contradicts that $q(\boldsymbol{h}) = 2q(\boldsymbol{h}_0)$. In every case, therefore, we have $-\boldsymbol{h} \in H_+$, hence $\boldsymbol{h} \in -H_+$. This proves $H_- \subseteq -H_+$, establishing the claim.

Claim 5. For every $\alpha > 0$, we have the following containments:

$$\alpha H_0 \subseteq H_0$$
$$\alpha H'_+ \subseteq H'_+$$
$$\alpha H_+ \subseteq H_+$$
$$\alpha H'_- \subseteq H'_-$$
$$\alpha H_- \subseteq H_-$$

Let $\boldsymbol{h} \in H_0$ be given, so $q(\boldsymbol{h}) = 0$. Then by (1.22), $q(\alpha\boldsymbol{h}) = \alpha q(\boldsymbol{h}) = 0$, so $\alpha\boldsymbol{h} \in H_0$. This proves $\alpha H_0 \subseteq H_0$.

Now let $\boldsymbol{h} \in H'_+$ be given. By (1.22), for every $\alpha > 0$ we have $q(\alpha\boldsymbol{h}) = \alpha q(\boldsymbol{h}) \neq 0$, so we may define a map $g : (0, \infty) \to R^n \setminus H_0$ by $g(\alpha) = \alpha\boldsymbol{h}$. Since g is continuous and its domain is connected, it follows that its image is also connected. Since $g(1) = \boldsymbol{h} \in H'_+$ and H'_+ is clopen in $\mathbb{R}^n \setminus H_0$, we deduce that the image of g is contained in H'_+. In other words, for every $\alpha > 0$ we have $\alpha\boldsymbol{h} \in H'_+ \subseteq H_+$, as desired. This proves that $\alpha H'_+ \subseteq H'_+$. Combining this with the containment $\alpha H_0 \subseteq H_0$ which was already shown, we also obtain $\alpha H_+ \subseteq H_+$.

The containment $\alpha H'_- \subseteq H_-$ may be proven in exactly the same way. Alternatively, we may derive it from what we have already shown, as follows: Let $\boldsymbol{h} \in H'_-$ and $\alpha > 0$ be given. We have either $\alpha\boldsymbol{h} \in H'_-$ or $\alpha\boldsymbol{h} \in H_+$. If $\alpha\boldsymbol{h} \in H_+$, then by what we just showed, it follows that $\boldsymbol{h} = \frac{1}{\alpha}(\alpha\boldsymbol{h}) \in H_+$, but this contradicts that $\boldsymbol{h} \in H'_-$, so we must have $\alpha\boldsymbol{h} \in H'_-$. This proves that $\alpha H'_- \subseteq H'_-$, and the containment $\alpha H_- \subseteq H_-$ immediately follows.

Claim 6. If $\boldsymbol{h}_1, \boldsymbol{h}_2 \in H_0$, then $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_0$.

Given $\boldsymbol{h}_1, \boldsymbol{h}_2 \in H_0$, we have $q(\boldsymbol{h}_1) = q(\boldsymbol{h}_2) = 0$, so by (1.20)

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2) \in \{q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2), |q(\boldsymbol{h}_1) - q(\boldsymbol{h}_2)|\} = \{0\}$$

Therefore, $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = 0$, i.e., $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_0$.

Claim 7. If $\boldsymbol{h}_1, \boldsymbol{h}_2 \in H_+$, then $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$ and $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_+$.

Let $\boldsymbol{h}_1, \boldsymbol{h}_2 \in H_+$ be given, so that $q(\boldsymbol{h}_1 + \boldsymbol{h}_0) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0)$ and $q(\boldsymbol{h}_2 + \boldsymbol{h}_0) = q(\boldsymbol{h}_2) + q(\boldsymbol{h}_0)$. Applying (1.20), we have

$$\begin{aligned} q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) &\in \{q(\boldsymbol{h}_1 + \boldsymbol{h}_0) + q(\boldsymbol{h}_2), |q(\boldsymbol{h}_1 + \boldsymbol{h}_0) - q(\boldsymbol{h}_2)|\} \\ &= \{q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0) + q(\boldsymbol{h}_2), |q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0) - q(\boldsymbol{h}_2)|\} \end{aligned} \tag{1.28}$$

and on the other hand

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) \in \{q(\boldsymbol{h}_1 + \boldsymbol{h}_2) + q(\boldsymbol{h}_0), |q(\boldsymbol{h}_1 + \boldsymbol{h}_2) - q(\boldsymbol{h}_0)|\} \tag{1.29}$$

Note that from (1.20) we have $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) \leq q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$. Suppose that the first case of (1.28) holds, i.e., that

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0) + q(\boldsymbol{h}_2). \tag{1.30}$$

Then we have

$$\begin{aligned} |q(\boldsymbol{h}_1 + \boldsymbol{h}_2) - q(\boldsymbol{h}_0)| &\leq q(\boldsymbol{h}_1 + \boldsymbol{h}_2) + q(\boldsymbol{h}_0) \\ &\leq q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2) + q(\boldsymbol{h}_0) \\ &= q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) \end{aligned}$$

If the first inequality here is an equality, then since $q(\boldsymbol{h}_0) > 0$ we must have $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = 0$, implying that $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_0 \subseteq H_+$. If the second inequality is an equality, then we have $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$. Therefore, if both inequalities are equalities, then both conclusions of the claim are satisfied. So assume that at least one of the inequalities is strict. Then the second case in (1.29) cannot occur, so we must have

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) = q(\boldsymbol{h}_1 + \boldsymbol{h}_2) + q(\boldsymbol{h}_0)$$

This means that $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_+$, and together with (1.30) it implies that $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$, so that again both conclusions of the claim are satisfied.

Without loss of generality, then, we may assume that the first case of (1.28) does *not* hold. We must therefore have

$$q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) = |q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0) - q(\boldsymbol{h}_2)| \tag{1.31}$$

Applying (1.20) once more, we have

$$\begin{aligned} q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) &\in \{q(\boldsymbol{h}_2 + \boldsymbol{h}_0) + q(\boldsymbol{h}_1), |q(\boldsymbol{h}_2 + \boldsymbol{h}_0) - q(\boldsymbol{h}_1)|\} \\ &= \{q(\boldsymbol{h}_2) + q(\boldsymbol{h}_0) + q(\boldsymbol{h}_1), |q(\boldsymbol{h}_2) + q(\boldsymbol{h}_0) - q(\boldsymbol{h}_1)|\} \end{aligned}$$

Since the first case here cannot hold, we must have the second case, which, combined with (1.31), implies that

$$|q(\boldsymbol{h}_1) + q(\boldsymbol{h}_0) - q(\boldsymbol{h}_2)| = |q(\boldsymbol{h}_2) + q(\boldsymbol{h}_0) - q(\boldsymbol{h}_1)|$$

Since $q(\boldsymbol{h}_0) \neq 0$, this implies that $q(\boldsymbol{h}_1) = q(\boldsymbol{h}_2)$.

We have shown then that if $q(\boldsymbol{h}_1) \neq q(\boldsymbol{h}_2)$ then the conclusions of the claim are satisfied. It only remains to handle the special case where $q(\boldsymbol{h}_1) = q(\boldsymbol{h}_2)$. In this case, if $q(\boldsymbol{h}_1) = q(\boldsymbol{h}_2) = 0$, then we are done by the previous claim. So assume $q(\boldsymbol{h}_1) = q(\boldsymbol{h}_2) \neq 0$. Given any positive $\alpha \neq 1$, by (1.22) we have $q(\alpha\boldsymbol{h}_2) = \alpha q(\boldsymbol{h}_2) = \alpha q(\boldsymbol{h}_1) \neq q(\boldsymbol{h}_1)$, and $\alpha\boldsymbol{h}_2 \in H_+$ since $\alpha H_+ \subseteq H_+$. Therefore, by the general case of the claim that we have already proven, we have $q(\boldsymbol{h}_1 + \alpha\boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\alpha\boldsymbol{h}_2)$. Since q is continuous, taking the limit as $\alpha \to 1$ gives $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$. Similarly, by the general case of the claim, for $\alpha \neq 1$ we have $\boldsymbol{h}_1 + \alpha\boldsymbol{h}_2 \in H_+$, i.e., $q(\boldsymbol{h}_1 + \alpha\boldsymbol{h}_2 + \boldsymbol{h}_0) = q(\boldsymbol{h}_1 + \alpha\boldsymbol{h}_2) + q(\boldsymbol{h}_0)$. Taking the limit as $\alpha \to 1$ gives $q(\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_0) = q(\boldsymbol{h}_1 + \boldsymbol{h}_2) + q(\boldsymbol{h}_0)$, i.e., $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_+$, completing the proof of the claim.

Claim 8. $L(\boldsymbol{h})$ is linear.

Applying the equality $-H_+ = H_-$ and equation (1.21) to the definition of $L(\boldsymbol{h})$, we have that

$$L(-\boldsymbol{h}) = -L(\boldsymbol{h}), \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^n \tag{1.32}$$

From the containments $\alpha H_+ \subseteq H_+$ and $\alpha H_- \subseteq H_-$ and the equation (1.22), we obtain

$$L(\alpha\boldsymbol{h}) = \alpha L(\boldsymbol{h}), \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^n, \alpha \geq 0 \tag{1.33}$$

Combining (1.32) and (1.33), we obtain

$$L(\alpha\boldsymbol{h}) = \alpha L(\boldsymbol{h}), \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^n, \alpha \in \mathbb{R} \tag{1.34}$$

Since $q(\boldsymbol{h}_1 + \boldsymbol{h}_2) = q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2)$ and $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_+$ for $\boldsymbol{h}_1, \boldsymbol{h}_2 \in H_+$, we have

$$L(\boldsymbol{h}_1 + \boldsymbol{h}_2) = L(\boldsymbol{h}_1) + L(\boldsymbol{h}_2), \quad \text{for all } \boldsymbol{h}_1, \boldsymbol{h}_2 \in H_+ \tag{1.35}$$

Negating both sides of this equality, applying (1.32), and using the equality $-H_+ = H_-$, it follows that

$$L(\boldsymbol{h}_1 + \boldsymbol{h}_2) = L(\boldsymbol{h}_1) + L(\boldsymbol{h}_2), \quad \text{for all } \boldsymbol{h}_1, \boldsymbol{h}_2 \in H_- \tag{1.36}$$

Finally, given $\boldsymbol{h}_1 \in H_+$ and $\boldsymbol{h}_2 \in H_-$, we have either $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_+$ or $\boldsymbol{h}_1 + \boldsymbol{h}_2 \in H_-$. In the former case, we apply (1.35) and then (1.32) to obtain

$$L(\boldsymbol{h}_1) = L((\boldsymbol{h}_1 + \boldsymbol{h}_2) + -\boldsymbol{h}_2)$$

$$= L(\boldsymbol{h}_1 + \boldsymbol{h}_2) + L(-\boldsymbol{h}_2)$$
$$= L(\boldsymbol{h}_1 + \boldsymbol{h}_2) - L(\boldsymbol{h}_2)$$

and it follows that $L(\boldsymbol{h}_1+\boldsymbol{h}_2) = L(\boldsymbol{h}_1)+L(\boldsymbol{h}_2)$. In the latter case, similarly we apply (1.36) and (1.32) to obtain

$$\begin{aligned} L(\boldsymbol{h}_2) &= L((\boldsymbol{h}_1 + \boldsymbol{h}_2) + -\boldsymbol{h}_1) \\ &= L(\boldsymbol{h}_1 + \boldsymbol{h}_2) + L(-\boldsymbol{h}_1) \\ &= L(\boldsymbol{h}_1 + \boldsymbol{h}_2) - L(\boldsymbol{h}_1) \end{aligned}$$

Therefore, we have

$$L(\boldsymbol{h}_1 + \boldsymbol{h}_2) = L(\boldsymbol{h}_1) + L(\boldsymbol{h}_2), \quad \text{for all } \boldsymbol{h}_1 \in H_+, \boldsymbol{h}_2 \in H_-$$

Combining this with (1.35) and (1.36), we conclude that

$$L(\boldsymbol{h}_1 + \boldsymbol{h}_2) = L(\boldsymbol{h}_1) + L(\boldsymbol{h}_2), \quad \text{for all } \boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^n$$

Together with (1.34) this proves that L is linear, completing the proof. □

Theorem 1.17. *Let $z(\boldsymbol{s})$ be a real-valued locally integrable function defined on $\mathbb{R}^n$. Assume that there is a function $\gamma : \mathbb{R}^n \to \mathbb{R}$ such that for every bounded Borel measurable set $A \subset \mathbb{R}^n$ with positive Lebesgue measure $\lambda(A) > 0$,*

$$\frac{1}{\lambda(A)} \int_A (z(\boldsymbol{s} + \boldsymbol{h}) - z(\boldsymbol{s}))^2 \, d\boldsymbol{s} = \gamma(\boldsymbol{h}) \tag{1.37}$$

Then $z(\boldsymbol{s})$ is almost everywhere equal to an affine transformation, i.e., there is a constant $c \in \mathbb{R}$ such that $z(\boldsymbol{s}) = c + L(\boldsymbol{s})$ for almost every $\boldsymbol{s}$, where $L(\alpha_1 \boldsymbol{s}_1 + \alpha_2 \boldsymbol{s}_2) = \alpha_1 L(\boldsymbol{s}_1) + \alpha_2 L(\boldsymbol{s}_2)$ for all $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^n$ and $\alpha_1, \alpha_2 \in \mathbb{R}$.

Note that the existence of the integral in (1.37) is part of the hypothesis; if we assumed that $z(\boldsymbol{s})$ were locally square-integrable, then this existence would already be insured.

Proof. Fix a vector $h \in \mathbb{R}^n$. The assumption (1.37) implies that

$$\int_A [(z(\boldsymbol{s} + \boldsymbol{h}) - z(\boldsymbol{s}))^2 - \gamma(\boldsymbol{h})] \, d\boldsymbol{s} = 0$$

for every bounded Borel measurable set $A \subseteq \mathbb{R}^n$. Applying Lemma 1.6 we deduce that $(z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}))^2 = \gamma(\boldsymbol{h})$ for all $\boldsymbol{s} \in \mathbb{R}^n \setminus N_{\boldsymbol{h}}$, where $N_{\boldsymbol{h}}$ is a null set depending on $\boldsymbol{h}$. Now set $q(\boldsymbol{h}) = \sqrt{\gamma(\boldsymbol{h})}$, so that

$$q(\boldsymbol{h}) = |z(\boldsymbol{s} + \boldsymbol{h}) - z(\boldsymbol{s})|, \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^n \setminus N_{\boldsymbol{h}}$$

If we now let Y be the unit cube $Y = [0,1]^n \subseteq \mathbb{R}^n$, then we have

$$q(\boldsymbol{h}) = \int_Y q(\boldsymbol{h})\ d\boldsymbol{s} = \int_Y |z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s})|\ d\boldsymbol{s}$$

By Lemma 1.15 it follows that $q(\boldsymbol{h})$ is a continuous function, with $q(\boldsymbol{0}) = 0$. We may write

$$z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) = \epsilon(\boldsymbol{s},\boldsymbol{h})q(\boldsymbol{h}), \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^n \setminus N_{\boldsymbol{h}}$$

for some measurable function $\epsilon : \mathbb{R}^n \times \mathbb{R}^n \to \{-1, 1\}$. Now, given $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^n$, there exists an $\boldsymbol{s}$ outside of the null set $N_{\boldsymbol{h}_1} \cup N_{\boldsymbol{h}_1+\boldsymbol{h}_2} \cup (N_{\boldsymbol{h}_2} - \boldsymbol{h}_1)$, and for such an $\boldsymbol{s}$ we have

$$\begin{aligned}
\epsilon(\boldsymbol{s},\boldsymbol{h}_1+\boldsymbol{h}_2)q(\boldsymbol{h}_1+\boldsymbol{h}_2) &= z(\boldsymbol{s}+\boldsymbol{h}_1+\boldsymbol{h}_2) - z(\boldsymbol{s}) \\
&= (z(\boldsymbol{s}+\boldsymbol{h}_1) - z(\boldsymbol{s})) + (z(\boldsymbol{s}+\boldsymbol{h}_1+\boldsymbol{h}_2) - z(\boldsymbol{s}+\boldsymbol{h}_1)) \\
&= \epsilon(\boldsymbol{s},\boldsymbol{h}_1)q(\boldsymbol{h}_1) + \epsilon(\boldsymbol{s}+\boldsymbol{h}_1,\boldsymbol{h}_2)q(\boldsymbol{h}_2)
\end{aligned}$$

It follows that for all $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^n$, we have $q(\boldsymbol{h}_1+\boldsymbol{h}_2) = \pm q(\boldsymbol{h}_1) \pm q(\boldsymbol{h}_2)$, and since q is nonnegative we have

$$q(\boldsymbol{h}_1+\boldsymbol{h}_2) \in \{q(\boldsymbol{h}_1) + q(\boldsymbol{h}_2), |q(\boldsymbol{h}_1) - q(\boldsymbol{h}_2)|\} \tag{1.38}$$

for all $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^n$. Now we may apply Lemma 1.16 to deduce that there exists a linear functional L with $q(\boldsymbol{h}) = |L(\boldsymbol{h})|$ for all $\boldsymbol{h} \in \mathbb{R}^n$. Choose a measurable function $\epsilon' : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that $\epsilon'(\boldsymbol{s},\boldsymbol{h}) = \epsilon(\boldsymbol{s},\boldsymbol{h}) \operatorname{sgn} L(\boldsymbol{h})$ for all $\boldsymbol{h} \notin \ker L$. Then

$$z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) = \epsilon'(\boldsymbol{s},\boldsymbol{h})L(\boldsymbol{h}), \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^n \setminus N_{\boldsymbol{h}},$$

and it follows that

$$\int_{\mathbb{R}^n} |z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) - \epsilon'(\boldsymbol{s},\boldsymbol{h})L(\boldsymbol{h})|\ d\boldsymbol{s} = 0 \tag{1.39}$$

for all $\boldsymbol{h} \in \mathbb{R}^n$. By Fubini's theorem we deduce that

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) - \epsilon'(\boldsymbol{s},\boldsymbol{h})L(\boldsymbol{h})|\ d\boldsymbol{h}\ d\boldsymbol{s} = 0$$

This implies that for almost all $\boldsymbol{s}$, we have $\int_{\mathbb{R}^n} |z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) - \epsilon'(\boldsymbol{s},\boldsymbol{h})L(\boldsymbol{h})|\ d\boldsymbol{h} = 0$. Therefore, there is a null set S_0 and null sets $M_{\boldsymbol{s}}$ for all $\boldsymbol{s} \notin S_0$ such that

$$z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}) = \epsilon'(\boldsymbol{s},\boldsymbol{h})L(\boldsymbol{h}), \quad \text{for all } \boldsymbol{s} \in \mathbb{R}^n \setminus S_0, \boldsymbol{h} \in \mathbb{R}^n \setminus M_{\boldsymbol{s}} \tag{1.40}$$

Choose some $\boldsymbol{s}_0 \in \mathbb{R}^n \setminus S_0$. If L is identically zero, then 1.40 implies that $z(\boldsymbol{s}_0+\boldsymbol{h}) = z(\boldsymbol{s}_0)$ for almost all $\boldsymbol{h}$, i.e., z is almost everywhere constant, in which case we are done. So assume L is not identically zero, in which case $\ker L$ is a proper subspace of $\mathbb{R}^n$ and hence is a null

set. Choose $\boldsymbol{h}_0 \in \mathbb{R}^n \setminus (\ker L \cup M_{\boldsymbol{s}_0} \cup (S_0 - \boldsymbol{s}_0))$. Then applying (1.40) three times, we have that for all $\boldsymbol{h} \in \mathbb{R}^n \setminus (M_{\boldsymbol{s}_0} \cup (M_{\boldsymbol{s}_0+\boldsymbol{h}_0} + \boldsymbol{h}_0))$,

$$\begin{aligned}\epsilon'(\boldsymbol{s}_0, \boldsymbol{h})L(\boldsymbol{h}) &= z(\boldsymbol{s}_0 + \boldsymbol{h}) - z(\boldsymbol{s}_0) \\ &= (z(\boldsymbol{s}_0 + \boldsymbol{h}_0) - z(\boldsymbol{s}_0)) + (z(\boldsymbol{s} + \boldsymbol{h}) - z(\boldsymbol{s}_0 + \boldsymbol{h}_0)) \\ &= \epsilon'(\boldsymbol{s}_0, \boldsymbol{h}_0)L(\boldsymbol{h}_0) + \epsilon'(\boldsymbol{s}_0 + \boldsymbol{h}_0, \boldsymbol{h} - \boldsymbol{h}_0)L(\boldsymbol{h} - \boldsymbol{h}_0)\end{aligned}$$

By the linearity of L, this becomes

$$(\epsilon'(\boldsymbol{s}_0, \boldsymbol{h}) - \epsilon'(\boldsymbol{s}_0 + \boldsymbol{h}_0, \boldsymbol{h} - \boldsymbol{h}_0))L(\boldsymbol{h}) = (\epsilon'(\boldsymbol{s}_0, \boldsymbol{h}_0) - \epsilon'(\boldsymbol{s}_0 + \boldsymbol{h}_0, \boldsymbol{h} - \boldsymbol{h}_0))L(\boldsymbol{h}_0) \tag{1.41}$$

If $\boldsymbol{h} \notin \ker L$ and $\boldsymbol{h} \notin \ker L + \boldsymbol{h}_0$, then $L(\boldsymbol{h})$ and $L(\boldsymbol{h}_0)$ are nonzero and distinct, so that (1.41) implies

$$\epsilon'(\boldsymbol{s}_0, \boldsymbol{h}) = \epsilon'(\boldsymbol{s}_0 + \boldsymbol{h}_0, \boldsymbol{h} - \boldsymbol{h}_0) = \epsilon'(\boldsymbol{s}_0, \boldsymbol{h}_0)$$

Therefore, we have

$$\epsilon'(\boldsymbol{s}_0, \boldsymbol{h}) = \epsilon'(\boldsymbol{s}_0, \boldsymbol{h}_0), \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^n \setminus (M_{\boldsymbol{s}_0} \cup (M_{\boldsymbol{s}_0+\boldsymbol{h}_0} + \boldsymbol{h}_0) \cup \ker L \cup (\ker L + \boldsymbol{h}_0))$$

Setting $\epsilon_0 = \epsilon'(\boldsymbol{s}_0, \boldsymbol{h}_0)$, (1.40) becomes

$$z(\boldsymbol{s}_0 + \boldsymbol{h}) - z(\boldsymbol{s}_0) = \epsilon_0 L(\boldsymbol{h})$$

for all $\boldsymbol{h} \in \mathbb{R}^n \setminus (M_{\boldsymbol{s}_0} \cup (M_{\boldsymbol{s}_0+\boldsymbol{h}_0} + \boldsymbol{h}_0) \cup \ker L \cup (\ker L + \boldsymbol{h}_0))$. Substituting $\boldsymbol{s} = \boldsymbol{s}_0 + \boldsymbol{h}$, we obtain that for all $\boldsymbol{s}$ outside of a null set,

$$\begin{aligned}z(\boldsymbol{s}) &= z(\boldsymbol{s}_0) + \epsilon_0 L(\boldsymbol{s} - \boldsymbol{s}_0) \\ &= z(\boldsymbol{s}_0) - \epsilon_0 L(\boldsymbol{s}_0) + \epsilon_0 L(\boldsymbol{s})\end{aligned}$$

The conclusion of the theorem then holds with $c = z(\boldsymbol{s}_0) - \epsilon_0 L(\boldsymbol{s}_0)$ and with $\epsilon_0 L$ in place of L. □

Example 1.1. Let $Z_0(s)$ be a stationary Gaussian process (over $\mathbb{R}$) with a semivariogram having finite range, e.g., with semivariogram

$$\gamma_0(h) = \begin{cases} |h|, & |h| \le 1 \\ 1, & \text{otherwise} \end{cases}$$

Let X be a Gaussian random variable with mean 0 and variance σ^2, independent of Z_0. Define $Z(s) = Z_0(s)X$. Then Z is a second-order stationary random field with semivariogram

$$\gamma(h) = E((Z(s+h) - Z(s))^2)$$

$$
\begin{aligned}
&= E((Z_0(s+h) - Z(s))^2 X^2) \\
&= E((Z_0(s+h) - Z(s))^2) E(X^2) \\
&= \sigma^2 \gamma_0(h)
\end{aligned}
$$

In particular, the semivariogram of Z has finite range. However, the empirical semivariogram is not a consistent estimator of the semivariogram in this case. Specifically, if we condition on X, then Z becomes m-dependent with semivariogram $\gamma_0(h)X^2$, so that under suitable assumptions on the sampling locations, the empirical semivariogram converges to $\gamma_0(h)X^2$, which is a non-degenerate random variable. This is contrary to statements in [10, p. 192] and illustrates the need for stronger assumptions such as mixing conditions (which of course Z would not satisfy here) in order to ensure the consistency of the empirical semivariogram.

CHAPTER 2

MIXING CONDITIONS

Mixing conditions on random fields are assumptions which assert, roughly speaking, that the dependence in the random field is localized, i.e., that the value of the random field at a given location can only depend strongly on the value of the random field at nearby locations. Mixing conditions are well-studied in the context of times series, stochastic processes, and spatial random fields on $\mathbb{Z}^n$; see the survey [4] and monograph [6] by Bradley. However, not as much has been written about mixing conditions for random fields on $\mathbb{R}^n$, the relevant setting for many applications in spatial statistics.

In §2.1, we introduce the five classical α-, ρ-, β-, ϕ-, and ψ-mixing coefficients and the associated mixing conditions on random fields. In §2.2, we describe the relationship among these mixing conditions in the case of Gaussian random fields. While widely-applicable sufficient conditions have been known for stationary Gaussian random fields on $\mathbb{R}$ and $\mathbb{Z}^n$ to be ρ-mixing [6, 23], this is apparently not the case for $\mathbb{R}^n$. To fill this void, in §2.3, we establish such sufficient conditions, in terms of the spectral density, and we give several examples. In §2.4, we extend these results to multivariate random fields, and more generally, to V-valued random fields on a locally compact abelian group G, where V is an arbitrary Hilbert space. Where results are stated without a reference or proof, a proof may be found at the end of the chapter.

2.1 Mixing coefficients and conditions

Fix a probability space $(\Omega, \mathcal{F}, P)$. By *random variable*, we mean a real-valued Borel-measurable function defined on Ω. Given a σ-algebra $\mathcal{A} \subseteq \mathcal{F}$, let $L^2(\mathcal{A})$ denote the set of $\mathcal{A}$-measurable random variables X such that $E(X^2) < \infty$. Given two random variables $X, Y \in L^2(\mathcal{A})$, for convenience we define their correlation to be zero if X or Y is degenerate:

$$\operatorname{Corr}(X,Y) = \begin{cases} 0, & \text{if } \operatorname{Var}(X) = 0 \text{ or } \operatorname{Var}(Y) = 0 \\ \dfrac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, & \text{otherwise} \end{cases}$$

Given two σ-algebras $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, the five classical mixing coefficients are defined as follows:

$$\alpha(\mathcal{A},\mathcal{B}) = \sup\{|P(A\cap B) - P(A)P(B)| : A\in\mathcal{A}, B\in\mathcal{B}\}$$
$$\rho(\mathcal{A},\mathcal{B}) = \sup\{|\operatorname{Corr}(X,Y)| : X\in L^2(\mathcal{A}), Y\in L^2(\mathcal{B})\}$$
$$\beta(\mathcal{A},\mathcal{B}) = \sup\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}|P(A_i\cap B_j) - P(A_i)P(B_j)|$$
$$\phi(\mathcal{A},\mathcal{B}) = \sup\{|P(B|A) - P(B)| : A\in\mathcal{A}, B\in\mathcal{B}, P(A)>0\}$$
$$\psi(\mathcal{A},\mathcal{B}) = \sup\left\{\left|\frac{P(A\cap B)}{P(A)P(B)} - 1\right| : A\in\mathcal{A}, B\in\mathcal{B}, P(A)>0, P(B)>0\right\}$$

where in the definition of β, the supremum is taken over all $I, J\in\mathbb{N}$ and all $A_1,\dots,A_I\in\mathcal{A}$ and $B_1,\dots,B_J\in\mathcal{B}$ such that $\{A_1,\dots,A_I\}$ and $\{B_1,\dots,B_J\}$ are partitions of Ω. These five mixing coefficients provide different ways of measuring the strength of dependence between two σ-algebras; in each case, the mixing coefficent is zero if $\mathcal{A}$ and $\mathcal{B}$ are independent and is positive otherwise. The following inequalities are well-known [4, Proposition 3.11]:

$$2\alpha(\mathcal{A},\mathcal{B}) \le \beta(\mathcal{A},\mathcal{B}) \le \phi(\mathcal{A},\mathcal{B}) \le \frac{1}{2}\psi(\mathcal{A},\mathcal{B}) \tag{2.1}$$
$$4\alpha(\mathcal{A},\mathcal{B}) \le \rho(\mathcal{A},\mathcal{B}) \le 2\sqrt{\phi(\mathcal{A},\mathcal{B})\phi(\mathcal{B},\mathcal{A})} \tag{2.2}$$

By *random field*, we mean a collection of random variables $Z(\boldsymbol{s})$ indexed by $s\in\mathbb{R}^n$. Given two subsets $A, B\subseteq\mathbb{R}^n$, we define the distance between them as

$$d(A,B) = \inf\{\|\boldsymbol{a}-\boldsymbol{b}\| : \boldsymbol{a}\in A, \boldsymbol{b}\in B\}$$

Given a random field $Z(\boldsymbol{s})$ on $\mathbb{R}^n$, the classical mixing coefficients on Z are defined as functions of a nonnegative real variable R:

$$\alpha(h) = \sup\alpha(\sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in A), \sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in B))$$
$$\rho(h) = \sup\rho(\sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in A), \sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in B))$$
$$\beta(h) = \sup\beta(\sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in A), \sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in B))$$
$$\phi(h) = \sup\phi(\sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in A), \sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in B))$$
$$\psi(h) = \sup\psi(\sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in A), \sigma(Z(\boldsymbol{s}) : \boldsymbol{s}\in B))$$

where the supremum is taken over all closed half-space $A, B\subseteq\mathbb{R}^n$ such that $d(A,B)>h$. The *interlaced mixing coefficients* $\alpha^*(h)$, $\rho^*(h)$, $\beta^*(h)$, $\phi^*(h)$, and $\psi^*(h)$ are defined in the same way, except that in the supremum, the subsets A and B are not restricted to be closed half-spaces but range over all subsets $A, B\subseteq\mathbb{R}^n$ such that $d(A,B)>h$,. The random field is said to be α-*mixing* if $\alpha(h)\to 0$ as $h\to\infty$. The terms α^*-*mixing*, ρ-*mixing*, ρ^*-*mixing*, etc. are defined similarly. It is clear that we have the following implications:

$$\alpha^*\text{-mixing} \Rightarrow \alpha\text{-mixing}$$

$$\rho^*\text{-mixing} \Rightarrow \rho\text{-mixing}$$
$$\beta^*\text{-mixing} \Rightarrow \beta\text{-mixing}$$
$$\phi^*\text{-mixing} \Rightarrow \phi\text{-mixing}$$
$$\psi^*\text{-mixing} \Rightarrow \psi\text{-mixing}$$

From the inequalities (2.1) and (2.2), we also have

$$\psi\text{-mixing} \Rightarrow \phi\text{-mixing} \Rightarrow \rho\text{-mixing} \Rightarrow \alpha\text{-mixing}$$
$$\phi\text{-mixing} \Rightarrow \beta\text{-mixing} \Rightarrow \alpha\text{-mixing}$$

In the same way, we have implications for the interlaced mixing conditions:

$$\psi^*\text{-mixing} \Rightarrow \phi^*\text{-mixing} \Rightarrow \rho^*\text{-mixing} \Rightarrow \alpha^*\text{-mixing}$$
$$\phi^*\text{-mixing} \Rightarrow \beta^*\text{-mixing} \Rightarrow \alpha^*\text{-mixing}$$

Thus, out of the ten conditions that we have introduced, ψ^*-mixing is the strongest, while α-mixing is the weakest.

If $Z(\boldsymbol{s})$ is a strictly stationary random field on $\mathbb{R}^n$, then some of the above mixing conditions are "too strong", in the sense that they degenerate into an assumption analogous to m-dependence. Specifically, if Z satisfies β^*-mixing, then there exists an $h > 0$ such that $\beta^*(h) = 0$, so that the dependence completely vanishes beyond distance h (see [3, Theorem 1(ii), Remark 3]). If $n \geq 2$, then the analagous statement is true even if Z is only assumed to be β-mixing [3, Theorem 1(i), Remark 3]. The same statements then apply to the even stronger assumptions of ϕ-mixing, ϕ^*-mixing, ψ-mixing, and ψ^*-mixing: any one of these conditions implies a kind of m-dependence in strictly stationary random fields on $\mathbb{R}^n$ if $n \geq 2$. In what follows, therefore, we will restrict attention to the conditions of α-, α^*-, ρ-, and ρ^*-mixing.

2.2 Mixing in Gaussian random fields

Let I and J be arbitrary sets, let X_i be a collection of random variables indexed by $i \in I$, and let Y_j be a collection of random variables indexed by $j \in J$. We define the *linear dependence coefficient* $r(X, Y)$ between the two collections of random variables as

$$r(X, Y) = \sup \operatorname{Corr}\left(\sum_{i \in I} a_i X_i, \sum_{j \in J} b_j Y_j\right)$$

where the supremum ranges over all choices of coefficients a_i and b_j such that only finitely many coefficients are nonzero. In this situation we also define mixing coefficients between the two collections:

$$\alpha(X, Y) = \alpha(\sigma(X_i : i \in I), \sigma(Y_j : j \in J))$$
$$\rho(X, Y) = \rho(\sigma(X_i : i \in I), \sigma(Y_j : j \in J))$$

Given a random field $Z(\boldsymbol{s})$, the *linear dependence coefficient* $r(h)$ of Z is defined by

$$r(h) = \sup r((Z(\boldsymbol{s}) : \boldsymbol{s} \in A), (Z(\boldsymbol{s}) : \boldsymbol{s} \in B))$$

for $h \geq 0$, where the supremum ranges over all closed half-spaces $A, B \subseteq \mathbb{R}^n$ with $d(A, B) \geq h$. The *interlaced linear dependence coefficient* $r^*(h)$ is defined in the same way, except that in the supremum, the subsets A and B are not restricted to be closed half-spaces but range over all subsets $A, B \subseteq \mathbb{R}^n$ such that $d(A, B) \geq h$. We say that a random field $Z(\boldsymbol{s})$ is r-mixing if $r(h) \to 0$ as $h \to \infty$, and r^*-mixing if $r^*(h) \to 0$.

If the two collections of random variables are jointly Gaussian, then it turns out that their ρ-mixing coefficient $\rho(X, Y)$ is identical to their linear dependence coefficient $r(X, Y)$:

Theorem 2.1 (Kolmogorov, Rozanov [23])**.** *Let I and J be arbitrary sets, let X_i be a collection of random variables indexed by $i \in I$, and let Y_j be a collection of random variables indexed by $j \in J$, and assume that these two collections together are jointly Gaussian. Then $\rho(X, Y) = r(X, Y)$.*

This provides a helpful step toward computing $\rho(X, Y)$, because the spaces of finite linear combinations of variables in X and Y are easier to deal with than the spaces of all Borel-measurable functions of variables in X and Y. This fact is used in order to prove the following result, which shows that for jointly Gaussian random variables, the ρ-mixing and α-mixing coefficients differ from each other by no more than a factor of 2π:

Theorem 2.2 (Kolmogorov, Rozanov [23])**.** *Let I and J be arbitrary sets, let X_i be a collection of random variables indexed by $i \in I$, and let Y_j be a collection of random variables indexed by $j \in J$, and assume that these two collections together are jointly Gaussian. Then*

$$4\alpha(X, Y) \leq \rho(X, Y) \leq 2\pi\alpha(X, Y)$$

In particular, given a Gaussian random field $Z(\boldsymbol{s})$, this implies that for all $h \geq 0$,

$$4\alpha(h) \leq \rho(h) \leq 2\pi\alpha(h)$$

$$4\alpha^*(h) \leq \rho^*(h) \leq 2\pi\alpha^*(h)$$

Theorem 2.1 implies that $\rho(h) = r(h)$ and $\rho^*(h) = r^*(h)$, so that in the case of a Gaussian random field, we have the following equivalences:

$$\alpha\text{-mixing} \Leftrightarrow r\text{-mixing} \Leftrightarrow \rho\text{-mixing}$$
$$\alpha^*\text{-mixing} \Leftrightarrow r^*\text{-mixing} \Leftrightarrow \rho^*\text{-mixing}$$

We mention in passing that if $Z(\boldsymbol{s})$ is any strictly stationary random field on $\mathbb{R}^n$, not necessarily Gaussian, then it is known that the equivalence between α^*- and ρ^*-mixing continues to hold, and for $n \geq 2$ the equivalence between α- and ρ-mixing conditions also continues to hold [5].

2.3 Spectral density and mixing

While mixing conditions are convenient to use as assumptions in proving asymptotic results for random fields, it can be highly nontrivial to show that mixing conditions are satisfied even for specific, simple examples. For second-order stationary random fields on $\mathbb{Z}^n$, there is a convenient sufficient condition for r^*-mixing that is widely applicable [4, Theorem 28.19(II)]:

Theorem 2.3. *Let $Z(\boldsymbol{s})$ be a second-order stationary random field on $\mathbb{Z}^n$, and assume that it has a continuous, positive spectral density $f(\boldsymbol{\omega})$. Then Z is r^*-mixing.*

We recall from the previous section that if Z is assumed to be a Gaussian random field, then the conclusion of r^*-mixing also implies α^*-mixing and ρ^*-mixing. In Theorem 2.3, the assumption that Z has a spectral density $f(\omega)$ means that the covariance function $C(\boldsymbol{h})$ of Z satisfies

$$C(\boldsymbol{h}) = \int_{[0,2\pi]^n} e^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{h}} f(\boldsymbol{\omega})\, d\boldsymbol{\omega}$$

If we also have $C \in L^1(\mathbb{Z}^n)$, then Fourier inversion can be used to express $f(\omega)$ explicitly:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^n} \sum_{\boldsymbol{h}\in\mathbb{Z}^n} e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{h}} C(\boldsymbol{h})$$

We illustrate this with the example of a time series with exponential covariance function, which arises in the context of $AR(1)$ models:

Theorem 2.4. *Let X_i be a second-order stationary time series with covariance function $C(h) = \sigma^2 e^{-a|h|}$ for some $a > 0$. Then X_i is r^*-mixing.*

Proof. Without loss of generality, assume $\sigma^2 = 1$. We compute the spectral density:

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} e^{-\mathrm{i}\omega h} e^{-a|h|} \\ &= \frac{1}{2\pi} \left(2 \operatorname{Re} \sum_{h=0}^{\infty} e^{-(\mathrm{i}\omega + a)h} - 1 \right) \\ &= \frac{1}{2\pi} \left(2 \operatorname{Re} \frac{1}{1 - e^{-\mathrm{i}\omega - a}} - 1 \right) \\ &= \frac{1}{2\pi} \left(2 \operatorname{Re} \frac{1 - e^{\mathrm{i}\omega - a}}{1 - 2e^{-a}\cos(\omega) + e^{-2a}} - 1 \right) \\ &= \frac{1}{2\pi} \left(\frac{2(1 - e^{-a}\cos(\omega))}{1 - 2e^{-a}\cos(\omega) + e^{-2a}} - 1 \right) \\ &= \frac{1}{2\pi} \left(\frac{1 - e^{-2a}}{1 - 2e^{-a}\cos(\omega) + e^{-2a}} \right) \end{aligned}$$

Here the denominator is bounded between $1 - 2e^{-a} + e^{-2a} = (1 - e^{-a})^2$ and $1 + 2e^{-a} + e^{-2a} = (1 + e^{-a})^2$, which implies that

$$f(\omega) \geq \frac{1}{2\pi} \left(\frac{1 - e^{-2a}}{(1 + e^{-a})^2} \right).$$

Now Theorem 2.3 gives the desired conclusion. □

Now we want to consider the case of random fields on $\mathbb{R}^n$. Unfortunately, the statement of Theorem 2.3 becomes false if $\mathbb{Z}^n$ is replaced by $\mathbb{R}^n$. Kolmogorov and Rozanov have given a formula for $\rho(h)$ in the case of a stationary Gaussian random field on $\mathbb{R}$ [23, Theorem 3]:

$$\rho(h) = \inf_{\phi} \operatorname*{ess\,sup}_{\omega \in \mathbb{R}} \left| 1 - \frac{e^{\mathrm{i}\omega h} \phi(\omega)}{f(\omega)} \right|$$

where the infimum is taken over all functions ϕ in the Hardy space H^1 on the complex lower half plane, and where $f(\omega)$ is the spectral density. In some cases, this formula can be used to show that a Gaussian process is ρ-mixing. However, there is an apparent absence of results in the literature providing sufficient conditions for a stationary Gaussian random field on $\mathbb{R}^n$ to be ρ-mixing when $n \geq 2$. To fill this void, we establish the following result:

Theorem 2.5. *Let $Z(\boldsymbol{s})$ be a second-order stationary random field over $\mathbb{R}^n$ with covariance function $C(\boldsymbol{h})$ and spectral density $f(\boldsymbol{\omega})$, and let $R > 0$ be given. Suppose there exists a bounded function $C_1(\boldsymbol{h})$ with support contained in B_R such that the Fourier transform $\hat{C}_1$ satisfies*

$$|\hat{C}_1(\boldsymbol{\omega}) - f(\boldsymbol{\omega})| \leq \epsilon f(\boldsymbol{\omega}) \tag{2.3}$$

for all $\boldsymbol{\omega} \in \mathbb{R}^n$*, where the Fourier transform is defined by*

$$\hat{C}_1(\boldsymbol{\omega}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{h}} C_1(\boldsymbol{h})\, d\boldsymbol{h}.$$

Then $r^*(R) \leq \epsilon$.

The assumption here requires that the spectral density $f(\boldsymbol{\omega})$ be well-approximated by a bandlimited function $\hat{C}_1$; this requirement can be understood, therefore, as a type of smoothness condition on the spectral density. As an application, we can show that Ornstein-Uhlenbeck processes are ρ^*-mixing:

Theorem 2.6. *Let* $Z(s)$ *be a stationary stochastic process (over* $\mathbb{R}$*) with an exponential covariance function* $C(h) = \sigma^2 e^{-|h|/h_0}$*. Then* $Z(s)$ *is* r^**-mixing, with* $r^*(R) \leq (\frac{R}{h_0} + 1)e^{-R/h_0}$.

Remark. Restricting the domain of Z to the integers, this provides an alternative proof of Theorem 2.4, with the added bonus of an explicit bound on the mixing coefficient.

Proof. Without loss of generality, assume $\sigma^2 = 1$ and $h_0 = 1$. We will apply Theorem 2.5 taking C_1 to be the even function given by $C_1(h) = (1 - e^{h-R})C(h)$ for $h \in [0, R]$ and $C_1(h) = 0$ for $h > R$. The intuition here is that we define $C_1(h)$ in such a way that it agrees closely with $C(h)$ for $h \ll R$ while continuously tapering off to 0 at R; if we had defined $C_1(h)$ by simply truncating $C(h)$ to zero for $h \geq R$, then this would have introduced a discontinuity in the function $C_1 - C$, with the result that the Fourier transform $\hat{C}_1 - f$ would decay at a rate of $1/\omega$, in which case (2.19) could not be satisfied.

We calculate

$$\begin{aligned}
\hat{C}_1(\omega) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-h\omega\mathrm{i}} C_1(h)\, dh \\
&= \frac{1}{2\pi} \int_0^R \cos(\omega h) C_1(h)\, dh \\
&= \frac{1}{2\pi} \int_0^R \cos(\omega h)(1 - e^{h-R})e^{-h}\, dh \\
&= \frac{1}{2\pi(1+\omega^2)} \left(1 - e^{-R}\left(\frac{\sin(\omega R)}{\omega} + \cos(\omega R)\right)\right)
\end{aligned}$$

On the other hand, a straightforward calculation shows that $f(\omega) = \frac{1}{2\pi(1+\omega^2)}$. Therefore, we have

$$\left|1 - \frac{\hat{C}_1(\omega)}{f(\omega)}\right| = \left|e^{-R}\left(\frac{\sin(\omega R)}{\omega} + \cos(\omega R)\right)\right| \leq (R+1)e^{-R}$$

Here the upper bound decays to zero as $R \to \infty$, which implies that (2.19) holds for arbitrarily small ϵ, as required. □

With some additional work, we can extend the result of Theorem 2.6 to $\mathbb{R}^2$:

Theorem 2.7. *Let $Z(\boldsymbol{s})$ be a random field over $\mathbb{R}^2$ with isotropic covariance function $C(h) = \sigma^2 e^{-h/h_0}$. Then Z is r^*-mixing. More precisely, the r^* mixing coefficients satisfy the bound*

$$r^*(R) \leq e^{-R/h_0}\left(2 + 140\pi^2\sqrt{5}\left(\frac{R}{h_0}\right)^2 + \frac{\pi^2}{2} + \frac{\pi^2\sqrt{10}}{3}\left(\frac{5R}{h_0} + 20\right)\right)$$

2.4 Random fields on LCA groups

In the previous section, we saw that different results were obtained when we dealt with random fields on $\mathbb{R}^n$, compared to random fields on $\mathbb{Z}^n$. In this section, we will develop methods that can simultaneously deal with both of these cases, as well as with hybrid cases such as random fields on $\mathbb{Z}^2 \times \mathbb{R}$, which would arise for instance when dealing with data at discrete points on a two-dimensional spatial grid over continuous time. In general, we will consider random fields on a locally compact abelian group G, which includes all the cases just mentioned as well as other cases. A locally compact abelian group, or LCA group, is an abelian group which is also a topological space, such that the group operations of multiplication and inversion are continuous. We will begin by reviewing some standard theory about LCA groups; proofs can be found in texts on harmonic analysis [35, 11].

Let $(G, +)$ be an LCA group. Then there is a positive Borel measure μ on G which is invariant, in the sense that $\mu(g+A) = \mu(A)$ for every measurable set $A \subseteq G$ and $g \in G$, and which is non-trivial, in the sense that $\mu(G) > 0$. Such a measure is called a *Haar measure* on G and is unique up a positive constant factor; we assume that a specific Haar measure μ has been chosen on G, and hereafter all integrals on G are understood to be taken with respect to this Haar measure.

Three important examples of LCA groups are the real numbers $\mathbb{R}$ under addition, the integers $\mathbb{Z}$ under addition, and the circle group $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ under multiplication. For canonical choices of Haar measures, on $\mathbb{R}$ we can take the Lebesgue measure, on $\mathbb{Z}$ we can take the counting measure, and on $\mathbb{T}$ we can take the measure induced by the map $x \mapsto e^{ix}$ from $[0, 2\pi]$ onto $\mathbb{T}$, where the measure on $[0, 2\pi]$ is the Lebesgue measure.

A continuous homomorphism $\chi : G \to \mathbb{T}$ is called a *character* of G. Under pointwise multiplication, the set of characters of G forms an abelian group, denoted $\hat{G}$, called the *dual group* of G. Endowing $\hat{G}$ with the compact-open topology, it can be shown that $\hat{G}$ is an LCA group, and it has a Haar measure, known as the *dual measure*, canonically associated with the given Haar measure on G. Unless otherwise specified, integrals over $\hat{G}$ are understood

to be taken with respect to the dual measure. For the examples in the previous paragraph, we have the following isomorphisms of LCA groups:

$$\hat{\mathbb{R}} \cong \mathbb{R}, \quad \hat{\mathbb{Z}} \cong \mathbb{T}, \quad \hat{\mathbb{T}} \cong \mathbb{Z}$$

In general, there is an isomorphism between an LCA group G and its double dual $\hat{\hat{G}}$, given by the Pontryagin duality map $\delta : G \to \hat{\hat{G}}$ where $\delta(g)(\chi) = \chi(g)$. Given a function $f \in L^1(G)$, its *Fourier transform* is a function $\hat{f} \in C_0(\hat{G})$ given by

$$\hat{f}(\chi) = \int f(g)\overline{\chi(g)}\, dg$$

A function $C : G \to \mathbb{R}$ is *positive semidefinite* if

$$\sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j C(g_i - g_j) \geq 0$$

for all $k \in \mathbb{N}$, $a_1, \ldots, a_k \in \mathbb{R}$, and $g_1, \ldots, g_k \in G$. By Bochner's theorem [35, 1.4.3], if C is an even positive semidefinite function, then there exists a positive, finite measure μ on $\hat{G}$ such that

$$C(g) = \int \chi(g)\, d\mu(\chi)$$

for all $g \in G$; we call μ the *spectral measure* of C.

Given a random field $Z(g)$ on G, we say that Z is second-order stationary if $E(Z(g))$ is a constant and $\mathrm{Cov}(Z(h), Z(h+g))$ depends only on g. In this case, the function

$$C(g) = \mathrm{Cov}(Z(h), Z(h+g))$$

is well-defined and is called the *covariance function* of Z. In this case, $C(g)$ is a positive semidefinite function and hence has a spectral measure μ. If μ is absolutely continuous with respect to the dual measure on $\hat{G}$, with density $f : G \to \mathbb{R}$, then we say that f is the *spectral density* of the random field.

Now we introduce a notion of α^*-, ρ^*-, and r^*-mixing coefficients for a random field $Z(g)$ on an LCA group G. Given a compact subset $K \subseteq G$, we define

$$\alpha^*(K) = \sup \alpha(\sigma(Z(g) : g \in A), \sigma(Z(g) : g \in B))$$
$$\rho^*(K) = \sup \rho(\sigma(Z(g) : g \in A), \sigma(Z(g) : g \in B))$$
$$r^*(K) = \sup r((Z(g) : g \in A), (Z(g) : g \in B))$$

where each supremum is taken over all pairs of subsets $A, B \subseteq G$ such that $(A+K) \cap B = \emptyset$. We note that if $G = \mathbb{R}^n$, then taking K to be a closed ball of radius h centered at the origin, we have $\alpha^*(K) = \alpha^*(h)$, $\rho^*(K) = \rho^*(h)$, and $r^*(K) = r^*(h)$, so in this way we recover the mixing coefficients as defined previously for random fields on $\mathbb{R}^n$.

Theorem 2.8. *Let $Z(g)$ be a second-order stationary random field over an LCA group G, with covariance function $C(g)$ and spectral density $f(\chi)$, and let a compact subset $K \subseteq G$ be given. If there exists a bounded function $C_1(h)$ with support contained in K such that the Fourier transform $\hat{C}_1$ satisfies*

$$|\hat{C}_1(\chi) - f(\chi)| \leq \epsilon f(\chi) \tag{2.4}$$

for all $\chi \in \hat{G}$, then $r^*(K) \leq \epsilon$.

From this general result, Theorems 2.3 and 2.5 follow as corollaries, by taking $G = \mathbb{Z}^n$ and $G = \mathbb{R}^n$, respectively.

2.5 Vector-valued random fields

In this section, we will extend the previous results to the more general setting of vector-valued random fields. We will need to deal with complex random variables, i.e., random elements of $\mathbb{C}$. Given two complex random variables X and Y, we define their covariance as the complex number given by

$$\mathrm{Cov}(X, Y) = E(X - E(X))\overline{(Y - E(Y))}$$

Given a real topological vector space V, we let V' denote its complex dual space, the space of complex-valued continuous linear functionals on V. A *random element* of V is understood to be defined with respect to the Borel σ-algebra under the weak*-topology, i.e., the σ-algebra generated by the linear functionals $\lambda \in V'$. In this setting, we define mixing coefficients as follows:

$$\alpha^*(K) = \sup \alpha(\sigma(\lambda(Z(g)) : g \in A, \lambda \in V'), \sigma(\lambda(Z(g)) : g \in B, \lambda \in V'))$$
$$\rho^*(K) = \sup \rho(\sigma(\lambda(Z(g)) : g \in A, \lambda \in V'), \sigma(\lambda(Z(g)) : g \in B, \lambda \in V'))$$
$$r^*(K) = \sup r((\lambda(Z(g)) : g \in A, \lambda \in V'), (\lambda(Z(g)) : g \in B, \lambda \in V'))$$

A random element X of V *has second moments* if $E[(\lambda(X))^2] < \infty$ for every continuous linear functional $\lambda \in V'$. Given two random elements X, Y of V, the *covariance* $\mathrm{Cov}(X, Y)$ of X and Y is defined as the sesquilinear form on V' given by

$$\mathrm{Cov}(X, Y)(\lambda_1, \lambda_2) = \mathrm{Cov}(\lambda_1(X), \lambda_2(Y))$$

Given an LCA group G, a V-valued *random field* on G is a collection of random elements of V, indexed by G. Such a random field is *second-order stationary* if the covariance between

$Z(g_1)$ and $Z(g_2)$ depends only on $g_2 - g_1$; i.e., there is a function C, called the *covariance function* of Z, such that

$$\mathrm{Cov}(Z(g_1), Z(g_2)) = C(g_2 - g_1)$$

We will say that C is *symmetric* if $C(-g) = C(g)$. We say that C is *integrable* if $C(g)(\lambda_1, \lambda_2)$ is an integrable function of g (with respect to the Haar measure on G) for every $\lambda_1, \lambda_2 \in V'$. In this case, the *spectral density* f is defined as the Fourier transform of C:

$$f(\chi) = \int C(g)\chi(g)\ dG(g)$$

where here the integral is understood to be defined componentwise, i.e.,

$$f(\chi)(\lambda_1, \lambda_2) = \int C(g)(\lambda_1, \lambda_2)\chi(g)\ dG(g)$$

Note that C is a collection of sesquilinear forms on V' indexed by G, while f is a collection of sesquilinear forms on V' indexed by $\hat{G}$. In this setting, Theorem 2.8 may be generalized as follows:

Theorem 2.9. *Let V be a topological vector space, and let $Z(g)$ be a second-order stationary V-valued random field over an LCA group G, with covariance function $C(g)$ and spectral density $f(\chi)$, and let a compact subset $K \subseteq G$ be given. If there exists an integrable function $C_1(h)$ with support contained in K such that the Fourier transform $\hat{C}_1$ satisfies*

$$|\hat{C}_1(\chi)(\lambda_1, \lambda_2) - f(\chi)(\lambda_1, \lambda_2)| \leq \epsilon\sqrt{f(\chi)(\lambda_1, \lambda_1)f(\chi)(\lambda_2, \lambda_2)} \tag{2.5}$$

for all $\lambda_1, \lambda_2 \in V'$ and all $\chi \in \hat{G}$, then $r^*(K) \leq \epsilon$.

Now we consider the special case where V is a real Hilbert space, with inner product $\langle \cdot, \cdot \rangle$. In this situation, the dual space V' becomes a complex Hilbert space in a natural way, and we will also write its inner product as $\langle \cdot, \cdot \rangle$. Furthermore, any bounded sesquilinear form $h(\cdot, \cdot)$ on V' may be represented by a bounded linear operator T_h on V', in the sense that

$$h(\lambda_1, \lambda_2) = \langle T_h\lambda_1, \lambda_2 \rangle$$

In particular, if we assume that the covariance $C(g)$ and spectral density $f(\chi)$ are bounded sesquilinear forms, then they may be represented by bounded linear operators $T_{C(g)}$ and $T_{f(\chi)}$, respectively, on V'. In this situation, we may put Theorem 2.9 into a more explicit form:

Theorem 2.10. *Let V be a Hilbert space, and let $Z(g)$ be a second-order stationary V-valued random field over an LCA group G, with symmetric covariance function $C(g)$ and positive definite spectral density $f(\chi)$, such that $f(\chi)$ is a bounded sesquilinear form for each χ. Let a compact subset $K \subseteq G$ be given, and let $C_1(g)$ be an integrable function with support contained in K, such that $\hat{C}_1(g)$ is a bounded sesquilinear form for all g. Then*

$$r^*(K) \leq \sup_{\chi \in \hat{G}} \operatorname{rad}\left(T^{-1}_{f(\chi)}(T_{\hat{C}_1(\chi)} - T_{f(\chi)})\right), \tag{2.6}$$

where rad *denotes the spectral radius of an operator.*

In the case where $V = \mathbb{R}^p$, given a sesquilinear form h on V', we let M_h denote the matrix representing T_h with respect to the standard basis $e_1^*, \ldots, e_p^*$ of $(\mathbb{R}^p)'$ (defined as the dual basis to the standard basis $e_1, \ldots, e_p$ of $\mathbb{R}^p$, i.e., by $e_i^*(e_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta function). In other words,

$$M_h = \big(\langle T_h(e_i^*), e_j^* \rangle\big) = \big(h(e_i^*, e_j^*)\big)$$

That is, M_h is the Gram matrix of the form h with respect to the basis $e_1^*, \ldots, e_p^*$. In particular, the covariance and spectral density forms are represented by matrices:

$$M_{C(g)} = \big(C(g)(e_i^*, e_j^*)\big)$$
$$M_{C(\chi)} = \big(f(\chi)(e_i^*, e_j^*)\big)$$

Equation (2.6) then becomes

$$r^*(K) \leq \sup_{\chi \in \hat{G}} \operatorname{rad}\left(M^{-1}_{f(\chi)}(M_{\hat{C}_1(\chi)} - M_{f(\chi)})\right),$$

As an application, we give the following example of a family of $\mathbb{R}^2$-valued random fields which are r^* mixing (hence, if we specialize to the case of Gaussian random fields, they are ρ^*-mixing):

Theorem 2.11. *Let $Z(s)$ be a second-order stationary $\mathbb{R}^2$-valued random field on $\mathbb{R}$ with covariance function given by*

$$M_{C(h)} = \begin{pmatrix} e^{-x} & ce^{-bx} \\ ce^{-bx} & e^{-ax} \end{pmatrix}$$

where $a > 0, b > 0$, and

$$c^2 < \min\left\{\frac{a}{b^2}, \frac{b^2}{a}, \frac{2a}{1+a^2}\right\} \tag{2.7}$$

Then $Z(s)$ is r^*-mixing. More precisely, if we set $K_1 = \min\{1, a, b\}$, then there is some constant L, depending only on a, b, and c, such that

$$r^*(R) \leq L(R+1)e^{-KR}$$

2.6 Proofs

We first turn our attention to Theorem 2.1, proven by Kolmogorov and Rozanov [23]. We will give a self-contained proof, including a fuller exposition of technical details which were left implicit in [23]. To begin, we make the following definition: given a σ-algebra $\mathcal{A} \subseteq \mathcal{F}$, we define the set of *standardized* $\mathcal{A}$-measurable random variables as

$$L_s^2(\mathcal{A}) = \{X \in L^2(\mathcal{A}) : E(X) = 0, \operatorname{Var}(X) = 1\}$$

Given collections of random variables $(X_i : i \in I)$ and $(Y_j : j \in J)$, it is straightforward to check that in the definition of the mixing coefficient $\rho(X, Y)$, we may restrict to standardized random variables; i.e.,

$$\rho(X, Y) = \sup\{\operatorname{Cov}(F, G) : F \in L_s^2(\sigma(X)), G \in L_s^2(\sigma(Y))\} \tag{2.8}$$

By [22, Lemma 1.13], every $\sigma(X)$-measurable random variable may be expressed as a measurable function of X, so we may also write

$$\rho(X, Y) = \sup\{\operatorname{Cov}(f(X), g(Y)) : f, g \text{ measurable}, f(X) \in L_s^2, g(Y) \in L_s^2(\sigma(Y))\} \tag{2.9}$$

The following result will allow us to derive an expression for the ρ-mixing coefficient in terms of conditional expectations:

Lemma 2.12 (Rényi [33])**.** *Given real, nondegenerate random variables X and Y with $X \in L_s^2$,*

$$\sqrt{\operatorname{Var}(E(X|Y))} = \sup\{|\operatorname{Cov}(X, G)| : G \in L_s^2(\sigma(Y))\}.$$

Proof. Given $G \in L_s^2(\sigma(Y))$, by basic properties of conditional expectation and the Cauchy-Schwarz inequality we have

$$\begin{aligned}
|\operatorname{Cov}(X, G)| &= |E(XG)| \\
&= |E(E(XG|Y))| \\
&= |E(E(X|Y)G)| \\
&\leq \sqrt{E([E(X|Y)]^2)E(G^2)} \\
&= \sqrt{\operatorname{Var}(E(X|Y))\operatorname{Var}(G)} \\
&= \sqrt{\operatorname{Var}(E(X|Y))}.
\end{aligned}$$

Taking the supremum over all $G \in L^2_s(\sigma(Y))$, we obtain

$$\sqrt{\operatorname{Var}(E(X|Y))} \geq \sup\{|\operatorname{Cov}(X,G)| : G \in L^2_s(\sigma(Y))\}$$

If $\sqrt{\operatorname{Var}(E(X|Y))} = 0$, then there is nothing more to prove, so assume $\sqrt{(\operatorname{Var}(E(X|Y))} \neq 0$. In the above calculation, equality occurs in the Cauchy-Schwarz inequality if G is a scalar multiple of $E(X|Y)$. In particular, if we set $G = E(X|Y)/\sqrt{\operatorname{Var}(E(X|Y))}$, then $G \in L^2_s(\sigma(Y))$ and the equality $|\operatorname{Cov}(X,G)| = \sqrt{\operatorname{Var}(E(X|Y))}$ holds, completing the proof. □

Lemma 2.13. *Given two real, nondegenerate random variables X and Y, we have*

$$\rho(X,Y) = \sup\{\sqrt{\operatorname{Var}(E(F|Y))} : F \in L^2_s(\sigma(X))\}.$$

Proof. Applying Lemma 2.12, by the definition of $\rho(X,Y)$ we have

$$\begin{aligned}\rho(X,Y) &= \sup\{|\operatorname{Cov}(F,G)| : F \in L^2_s(\sigma(X)), G \in L^2_s(\sigma(Y))\} \\ &= \sup\{\sup\{|\operatorname{Cov}(F,G)| : G \in L^2_s(\sigma(Y))\} : F \in L^2_s(\sigma(X))\} \\ &= \sup\{\sqrt{\operatorname{Var}(E(F|Y))} : F \in L^2_s(\sigma(X))\}\end{aligned}$$

□

Lemma 2.14. *Let X, Y, and Z be random variables with $X, Y \in L^1$ and $(X,Z) \stackrel{d}{=} (Y,Z)$. Then $E(X|Z) = E(Y|Z)$ almost surely.*

Proof. For any measurable set $A \subseteq \mathbb{R}$, we have $E(E(X|Z)1_A(Z)) = E(X1_A(Z))$ and $E(E(Y|Z)1_A(Z)) = E(Y1_A(Z))$ by the definition of conditional expectation. Since $(X,Z) \stackrel{d}{=} (Y,Z)$, we have $E(X1_A(Z)) = E(Y1_A(Z))$, hence $E(E(X|Z)1_A(Z)) = E(E(Y|Z)1_A(Z))$, and the conclusion follows by the uniqueness of conditional expectation. □

The following special case of Theorem 2.1 was shown by Gebelein [15] and Lancaster [28]; we will give a proof based on that of Yu [41]:

Theorem 2.15. *Let X, Y be jointly Gaussian, nondegenerate random variables. Then $\rho(X,Y) = |\operatorname{Corr}(X,Y)|$.*

Proof. Without loss of generality, we may assume $X, Y \in L^2_s$, and in this case, the joint distribution of X and Y is completely determined by $\operatorname{Cov}(X,Y)$ (see Theorem 1.5), so we may define a function $\rho(r)$ on $[-1,1]$ by

$$\rho(r) = \rho(X,Y), \quad \text{where } \operatorname{Cov}(X,Y) = r$$

What we must prove is that $\rho(r) = |r|$. If $\mathrm{Cov}(X,Y) = r$, then $\mathrm{Cov}(-X,Y) = -r$, and $L_s^2(\sigma(X)) = L_s^2(\sigma(-X))$, hence $\rho(-X,Y) = \rho(X,Y)$; i.e., $\rho(-r) = \rho(r)$. Therefore, it suffices to show that $\rho(r) = r$ for $r \in [0,1]$. We now proceed by proving several claims:

Claim 1. For all $r \in [0,1]$, $\rho(r) \geq r$.

Given $r \in [0,1]$, choose jointly Gaussian random variables $X, Y \in L_s^2$ with $\mathrm{Cov}(X,Y) = r$. Then by the definitions of $\rho(r)$ and $\rho(X,Y)$, we have $\rho(r) = \rho(X,Y) \geq |\mathrm{Cov}(X,Y)| = r$.

Claim 2. $\rho(2^{-1/2}) = 2^{-1/2}$.

Let X and Y be independent standard Gaussian random variables. Then $\mathrm{Corr}(X, X+Y) = 2^{-1/2}$, so that $\rho(2^{-1/2}) = \rho(X, X+Y)$. Given measurable functions f and h such that $f(X), h(X+Y) \in L_s^2(\Omega)$, we have $(f(X), h(X+Y)) \stackrel{d}{=} (f(Y), h(X+Y))$ so that $f(Y) \in L_s^2(\sigma(Y))$ and $\mathrm{Cov}(f(X), h(X+Y)) = \mathrm{Cov}(f(Y), h(X+Y))$. If we set $r = \mathrm{Cov}(f(X), h(X+Y))$, then the covariance matrix of the random vector $(f(X), f(Y), h(X+Y))$ may be written

$$\mathrm{Var}\begin{pmatrix} f(X) \\ f(Y) \\ h(X+Y) \end{pmatrix} = \begin{pmatrix} 1 & 0 & r \\ 0 & 1 & r \\ r & r & 1 \end{pmatrix}$$

Since a covariance matrix must be positive semidefinite, in particular the determinant must be nonnegative, so that $1 - 2r^2 \geq 0$. Therefore, $\mathrm{Cov}(f(X), h(X+Y)) = r \leq 2^{-1/2}$. By (2.9) this proves that $\rho(2^{-1/2}) = \rho(X, X+Y) \leq 2^{-1/2}$. But we also have $\rho(2^{-1/2}) = \rho(X, X+Y) \geq \mathrm{Cov}(X, X+Y) = 2^{-1/2}$, so the equality $\rho(2^{-1/2}) = 2^{-1/2}$ holds, as claimed.

Claim 3. For all $r_1, r_2 \in [0,1]$, we have $\rho(r_1 r_2) \leq \rho(r_1)\rho(r_2)$, with equality if $r_1 = r_2$.

Given $r_1, r_2 \in [0,1]$, let N_1, N_2, and N_3 be independent standard Gaussian random variables, and define

$$\begin{aligned} X &= r_1 N_1 + \sqrt{1 - r_1^2} N_2 \\ Y &= r_2 N_1 + \sqrt{1 - r_2^2} N_3 \\ Z &= N_1 \end{aligned}$$

Note that $X, Y, Z \in L_s^2$ and that X and Y are conditionally independent given Z. Therefore, given $F \in L_s^2(\sigma(X))$ and $G \in L_s^2(\sigma(Y))$, we also have that F and G are conditionally independent given Z, and hence by the Cauchy-Schwarz inequality and Lemma 2.13,

$$\begin{aligned} \mathrm{Cov}(F,G) &= E(FG) \\ &= E(E(FG|Z)) \\ &= E(E(F|Z)E(G|Z)) \end{aligned}$$

$$\begin{aligned}
&= \mathrm{Cov}(E(F|Z), E(G|Z)) \\
&\leq \sqrt{\mathrm{Var}(E(F|Z))\,\mathrm{Var}(E(G|Z))} \\
&\leq \rho(X,Z)\rho(Y,Z) \\
&= \rho(r_1)\rho(r_2)
\end{aligned}$$

Taking the supremum over all such F and G gives $\rho(X,Y) \leq \rho(r_1)\rho(r_2)$, and since $\mathrm{Cov}(X,Y) = r_1 r_2$, this gives $\rho(r_1 r_2) \leq \rho(r_1)\rho(r_2)$, which proves the first part of the claim.

Now, if $r_1 = r_2$, then $(X,Z) \stackrel{d}{=} (Y,Z)$, so for any measurable function f with $f(X) \in L_s^2$ (hence $f(Y) \in L_s^2$) we have $(f(X),Z) \stackrel{d}{=} (f(Y),Z)$, and applying Lemma 2.14 gives

$$\begin{aligned}
\rho(X,Y) &\geq \mathrm{Cov}(f(X), f(Y)) \\
&= E(E(f(X)|Z)E(f(Y)|Z)) \\
&= E([E(f(X)|Z)]^2) \\
&= \mathrm{Var}(E(f(X)|Z))
\end{aligned}$$

Taking the supremum over all measurable functions f with $f(X) \in L_s^2$ and applying Lemma 2.13, this becomes

$$\rho(X,Y) \geq (\rho(X,Z))^2.$$

Since the reverse inequality has already been proven, it follows that equality holds, and in other words, $\rho(r_1^2) = \rho(r_1)^2$ as claimed.

Claim 4. ρ is an increasing function; i.e., for $0 \leq r_1 < r_2 \leq 1$, we have $\rho(r_1) \leq \rho(r_2)$.
Given $r_1 < r_2$, applying Claim 3 with r_1/r_2 in place of r_1 gives

$$\rho(r_1) = \rho\left((r_1/r_2)r_2\right) \leq \rho\left(r_1/r_2\right)\rho(r_2) \leq \rho(r_2).$$

Claim 5. For all integers $k \geq 1$, $\rho(2^{-1/2^k}) = 2^{-1/2^k}$.
We use induction on k. For $k = 1$, this is Claim 2. Assuming the claim is true for a certain k, we have $\rho(2^{-1/2^k}) = 2^{-1/2^k}$, and applying Claim 3 with $r_1 = r_2 = \rho(2^{-1/2^{k+1}}) = 2^{-1/2^{k+1}}$ gives

$$\rho\left(2^{-1/2^{k+1}}\right)^2 = \rho\left(\left(2^{-1/2^{k+1}}\right)^2\right) = \rho\left(2^{-1/2^k}\right) = 2^{-1/2^k}$$

Taking square roots, this gives $\rho(2^{-1/2^{k+1}}) = 2^{-1/2^{k+1}}$, as required.

Claim 6. For all integers $m \geq 1$ and $k \geq 1$, $\rho(2^{-m/2^k}) = 2^{-m/2^k}$.
We use induction on m. For $m = 1$, this is Claim 5. Assuming the claim is true for a certain m, by Claims 3 and 5 we have

$$\rho(2^{-(m+1)/2^k}) = \rho(2^{-1/2^k}2^{-m/2^k})$$

$$\begin{aligned}
&\leq \rho(2^{-1/2^k})\rho(2^{-m/2^k}) \\
&= 2^{-1/2^k}2^{-m/2^k} \\
&= 2^{-(m+1)/2^k}
\end{aligned}$$

On the other hand, we have the reverse inequality by Claim 1.

Claim 7. For all $0 \leq r \leq 1$, $\rho(r) = r$. Taking X and Y to be independent standard Gaussian random variables, we have $X, Y \in L^2_s$ and $\mathrm{Cov}(X, Y) = 0$ and also $\rho(X, Y) = 0$, so $\rho(0) = 0$. Taking X to be standard Gaussian, we have $\mathrm{Cov}(X, X) = 1$, so $\rho(1) = \rho(X, X) \geq 1$, which implies $\rho(1) = 1$.

Now, the set $\{-m/2^k : m, k \in \mathbb{N}\}$ is dense in $(-\infty, 0)$, and since the map $x \to 2^x$ is a homeomorphism from $(-\infty, 0)$ onto the $(0, 1)$, the image $Q = \{2^{-m/2^k} : m, k \in \mathbb{N}\}$ is dense in $(0, 1)$. Therefore, given r with $0 < r < 1$, for sufficiently small ϵ, there exist $r_1, r_2 \in Q$ with $r - \epsilon \leq r_1 \leq r \leq r_2 \leq r + \epsilon$. By Claim 4 this implies

$$r - \epsilon \leq r_1 = \rho(r_1) \leq \rho(r) \leq \rho(r_2) = r_2 \leq r + \epsilon$$

Taking $\epsilon \to 0$, this implies $\rho(r) = r$. □

Lemma 2.16. *Let $\mathcal{A}$ be an algebra of events (i.e., a subalgebra of the σ-algebra $\mathcal{F}$). For every event $B \in \sigma(\mathcal{A})$, given $\epsilon > 0$, there exists some $A \in \mathcal{A}$ with $P(A \triangle B) < \epsilon$.*

Proof. Let $\mathcal{B}$ denote the collection of events $B \in \sigma(\mathcal{A})$ satisfying the statement of the lemma. Clearly $\mathcal{A} \subseteq \mathcal{B}$, so if we can show that $\mathcal{B}$ is a σ-algebra, then it will follow that $\mathcal{B} = \sigma(\mathcal{A})$, as desired.

Let $B \in \mathcal{B}$ and $\epsilon > 0$ be given. We have $P(A \triangle B) < \epsilon$ for some $A \in \mathcal{A}$, hence $P(A^c \triangle B^c) = P(A \triangle B) < \epsilon$, and $A^c \in \mathcal{A}$, so that $B^c \in \mathcal{B}$. Therefore, B is closed under taking complements.

Now let $B_1, B_2, \cdots \in \mathcal{B}$ and $\epsilon > 0$ be given. Note that $P\left(\bigcup_{i=1}^k B_i\right)$ is an increasing sequence in k, with limit $P\left(\bigcup_{i=1}^\infty B_i\right)$, so we may choose k such that

$$P\left(\bigcup_{i=1}^{\infty} B_i\right) - P\left(\bigcup_{i=1}^{k} B_i\right) < \epsilon/2.$$

Now for each B_i, there exists some $A_i \in \mathcal{A}$ with $P(A_i \triangle B_i) < \frac{\epsilon}{2k}$, and we have

$$\begin{aligned}
P\left(\left(\bigcup_{i=1}^{k} A_i\right) \triangle \left(\bigcup_{i=1}^{k} B_i\right)\right) &\leq P\left(\bigcup_{i=1}^{k} (A_i \triangle B_i)\right) \\
&\leq \sum_{i=1}^{k} P(A_i \triangle B_i)
\end{aligned}$$

$$\leq \sum_{i=1}^{k} \frac{\epsilon}{2k}$$
$$= \epsilon/2$$

Therefore,

$$\begin{aligned}
&P\left(\left(\bigcup_{i=1}^{k} A_i\right) \triangle \left(\bigcup_{i=1}^{\infty} B_i\right)\right) \\
&\leq P\left(\left(\bigcup_{i=1}^{k} A_i\right) \triangle \left(\bigcup_{i=1}^{k} B_i\right)\right) + P\left(\left(\bigcup_{i=1}^{k} B_i\right) \triangle \left(\bigcup_{i=1}^{\infty} B_i\right)\right) \\
&\leq P\left(\left(\bigcup_{i=1}^{k} A_i\right) \triangle \left(\bigcup_{i=1}^{k} B_i\right)\right) + P\left(\bigcup_{i=1}^{\infty} B_i\right) - P\left(\bigcup_{i=1}^{k} B_i\right) \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon,
\end{aligned}$$

which proves that $\bigcup_{i=1}^{\infty} B_i \in \mathcal{B}$; i.e., $\mathcal{B}$ is closed under countable unions. Thus $\mathcal{B}$ is a σ-algebra, completing the proof. □

Let I be an arbitrary set, and suppose we are given a collection of random variables X_i indexed by $i \in I$. Given a subset $J \subseteq I$, we let X_J denote the subcollection of random variables X_j indexed by $j \in J$.

Lemma 2.17. *Let an arbitrary set I be given, and let X_i be a collection of random variables indexed by $i \in I$. Given a random variable $Y \in L^2(\sigma(X))$ and $\epsilon > 0$, there is some finite subset $I_0 \subseteq I$ and some random variable $Z \in L^2(\sigma(X_{I_0}))$ such that $\|Y - Z\| < \epsilon$, where $\|\cdot\|$ denotes the norm in L^2, i.e. $\sqrt{E[(\cdot)^2]}$.*

Proof. Define

$$\mathcal{A} = \bigcup \{\sigma(X_{I_0}) : I_0 \subseteq I, |I_0| < \infty\}$$

and note that $\mathcal{A}$ is an algebra: for if $A \in \mathcal{A}$, then $A \in \sigma(X_{I_0})$ for some finite subset $I_0 \subseteq I$, hence $A^c \in \sigma(X_{I_0}) \subseteq \mathcal{A}$, and if $A_1, A_2 \in \mathcal{A}$, then $A_1 \in \sigma(X_{I_1})$ and $A_2 \in \sigma(X_{I_2})$ for some finite sets $I_1, I_2 \subseteq \mathcal{A}$, hence $A_1 \cup A_2 \in \sigma(X_{I_1 \cup I_2}) \subseteq \mathcal{A}$.

Now, since simple functions are dense in $L^2(\sigma(X))$ (see [34, Theorem 3.13]), there is a random variable Y_0 with $\|Y - Y_0\| < \epsilon/2$, where Y_0 is of the form

$$Y_0 = \sum_{i=1}^{k} y_i 1_{B_i}$$

where $B_1, \ldots, B_k \in \sigma(X)$. If $y_1, \ldots, y_k$ are all zero, then we may take $Z = Y_0 = 0$, and we are done. Otherwise, applying Lemma 2.16, we may obtain events $A_1, \ldots, A_k \in \mathcal{A}$ with $P(A_i \triangle B_i) < \left(\frac{1}{2}\epsilon / \sum_{i=1}^k |y_i|\right)^2$. Define

$$Z = \sum_{i=1}^k y_i 1_{A_i}$$

Since $A_i \in \mathcal{A}$, for each i we have $A_i \in \sigma(X_{I_i})$ for some finite subset $I_i \subseteq I$. Setting $I_0 = \bigcup_{i=1}^k I_i$, we then have $A_i \in \sigma(X_{I_0})$ for all i, so that Z is $\sigma(X_{I_0})$-measurable. Now, we have

$$\begin{aligned}
\|Y_0 - Z\| &= \left\| \sum_{i=1}^k y_i (1_{B_i} - 1_{A_i}) \right\| \\
&\leq \sum_{i=1}^k |y_i| \|1_{B_i} - 1_{A_i}\| \\
&= \sum_{i=1}^k |y_i| \sqrt{P(A_i \triangle B_i)} \\
&\leq \frac{1}{2}\epsilon
\end{aligned}$$

Therefore, $Z \in L^2(\sigma(X_{I_0}))$ and satisfies

$$\|Y - Z\| \leq \|Y - Y_0\| + \|Y_0 - Z\| \leq \epsilon$$

as required. □

Lemma 2.18. *Given arbitrary sets I and J, and a collection of random variables X_i indexed by $i \in I$ and a collection of random variables Y_j indexed by $j \in J$, we have*

$$\rho(X, Y) = \sup\{\rho(X_{I_0}, Y_{J_0}) : I_0 \subseteq I, J_0 \subseteq J, |I_0| < \infty, |J_0| < \infty\} \tag{2.10}$$

Proof. Let $\rho_f(X, Y)$ be the right side of (2.10). The inequality $\rho(X, Y) \geq \rho_f(X, Y)$ is clear from (2.8), since $L^2_s(X_{I_0}) \subseteq L^2_s(X)$ and $L^2_s(Y_{J_0}) \subseteq L^2_s(Y)$ for every $I_0 \subseteq I$ and $J_0 \subseteq J$. Let $\epsilon > 0$ be given. Then there exist $F \in L^2_s(\sigma(X))$ and $G \in L^s_2(\sigma(Y))$ with $\rho(X, Y) \leq \mathrm{Cov}(F, G) + \epsilon$. By Lemma 2.17, for every $\epsilon_1 > 0$, there exist finite subsets $I_0 \subseteq I$ and $J_0 \subseteq J$ and random variables $F_0 \in \sigma(X_{I_0}), G_0 \in \sigma(Y_{J_0})$ such that $\|F - F_0\| < \epsilon_1$ and $\|G - G_0\| < \epsilon_1$. Therefore, $\|F_0\| < 1 + \epsilon_1$, and by the Cauchy-Schwarz inequality

$$\begin{aligned}
E(F_0 G_0) &= E(FG) - E((F - F_0)G) - E(F_0(G - G_0)) \\
&\geq \rho(X, Y) - \epsilon - \|F - F_0\|\|G\| - \|F_0\|\|G - G_0\|
\end{aligned}$$

$$\begin{aligned} &\geq \rho(X,Y) - \epsilon - \epsilon_1 - (1+\epsilon_1)\epsilon_1 \\ &= \rho(X,Y) - \epsilon - 2\epsilon_1 - \epsilon_1^2 \end{aligned}$$

Again, by the Cauchy-Schwarz inequality, we have $|E(F_0)| = |E(F - F_0)| \leq \|F - F_0\| \leq \epsilon_1$ and likewise $|E(G_0)| \leq \epsilon_1$, and hence

$$\begin{aligned} \mathrm{Cov}(F_0, G_0) &= E(F_0 G_0) - E(F_0)E(G_0) \\ &\geq \rho(X,Y) - \epsilon - 2\epsilon_1 - \epsilon_1^2 - \epsilon_1^2 \\ &\geq \rho(X,Y) - \epsilon - 2\epsilon_1 - 2\epsilon_1^2 \end{aligned}$$

From this we obtain

$$\begin{aligned} \rho_f(X,Y) &\geq \mathrm{Corr}(F_0, G_0) \\ &= \frac{\mathrm{Cov}(F_0, G_0)}{\sqrt{\mathrm{Var}(F_0)\,\mathrm{Var}(G_0)}} \\ &\geq \frac{\rho(X,Y) - \epsilon - 2\epsilon_1 - 2\epsilon_1^2}{(1+\epsilon_1)^2} \end{aligned}$$

Taking the limit as $\epsilon_1 \to 0$ and $\epsilon \to 0$, we conclude that $\rho_f(X,Y) \geq \rho(X,Y)$, as required. □

Theorem 2.19. *Let V be a real Hilbert space, and let U and W be finite-dimensional subspaces. Then there exists an orthonormal basis $u_1, \ldots, u_m$ of U and an orthonormal basis $w_1, \ldots, w_n$ of W such that $\langle u_i, v_j \rangle = 0$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$ with $i \neq j$.*

Proof. Given $u \in U$, the map $w \mapsto \langle u, w \rangle$ is a linear functional on W, so by the Riesz representation theorem there is a unique vector $\phi u \in W$ such that

$$\langle \phi u, w \rangle = \langle u, w \rangle$$

for all $w \in W$, and it is easy to check that the function $\phi : U \to W$ so determined is a linear mapping. Similarly, there is is a linear mapping $\psi : W \to U$ such that

$$\langle \psi w, u \rangle = \langle w, u \rangle$$

for all $u \in U$ and all $w \in W$. The composition $\psi\phi : U \to U$ is positive semidefinite, since for all $u \in U$,

$$\langle \psi\phi u, u \rangle = \langle \phi u, u \rangle = \langle u, \phi u \rangle = \langle \phi u, \phi u \rangle \geq 0 \tag{2.11}$$

Therefore, by the spectral theorem, the eigenvalues of $\psi\phi$ are real and $\psi\phi$ is diagonalizable, so that $\psi\phi$ has a nonzero eigenvalue unless $\psi\phi = 0$. But if $\psi\phi = 0$, then (2.11) shows that also $\phi = 0$, so that $\langle u, w \rangle = \langle \phi u, w \rangle = 0$ for all $u \in U$ and $w \in W$. In this case, the

statement of theorem is satisfied if we take $u_1, \ldots, u_m$ and $w_1, \ldots, w_n$ to be any orthonormal bases of U and W, respectively (such bases exist by the Gram-Schmidt process).

So we may assume that $\psi\phi$ has a non-zero eigenvalue $\lambda \in \mathbb{R}$ with eigenvector $u_0 \in U$. Let $w_0 = \phi u_0$, and note that $w_0 \neq 0$, since $0 \neq \lambda u_0 = \psi\phi u_0 = \psi w_0$. Now, setting $U_1 = U \cap \mathrm{Span}(u_0)^\perp$ and $W_1 = W \cap \mathrm{Span}(w_0)^\perp$, we may express U and W as orthogonal direct sums:

$$U = \mathrm{Span}(u_0) \oplus U_1$$
$$W = \mathrm{Span}(w_0) \oplus W_1$$

where $\dim(U_1) = \dim(U) - 1$ and $\dim(W_1) = \dim(W) - 1$. By induction, we may find an orthonormal basis $u_1, \ldots, u_m$ of U_1 and an orthonormal basis $w_1, \ldots, w_n$ of W_1 satisfying the statement of the theorem with U_1 in place of U and W_1 in place of W. For every $w \in W_1$, we have

$$0 = \langle w_0, w \rangle = \langle \phi u_0, w \rangle = \langle u_0, w \rangle$$

In particular, $\langle u_0, w_j \rangle = 0$ for $j = 1, \ldots, n$. Likewise, for every $u \in U_1$, we have

$$0 = \langle u_0, \lambda u \rangle = \langle \lambda u_0, u \rangle = \langle \psi\phi u_0, u \rangle = \langle \phi u_0, u \rangle = \langle w_0, u \rangle$$

In particular, $\langle u_i, w_0 \rangle = 0$ for $i = 1, \ldots, m$. Therefore, the orthonormal bases $\frac{u_0}{\|u_0\|}, u_1, \ldots, u_m$ and $\frac{w_0}{\|w_0\|}, w_1, \ldots, w_n$ satisfy the statement of the theorem. □

Lemma 2.20. *Let X_1, X_2 and Y_1 be random elements of measurable spaces A_1, A_2, and A_3, respectively, such that (X_1, Y_1) is independent of X_2, and let $F \in L^1$. Then*

$$E(E(F|X_1, X_2)|Y_1) = E(E(F|X_1)|Y_1)$$

Proof. Since (X_1, Y_1) is independent of X_2, it follows that X_2 is conditionally independent of Y_1 given X_1 and hence that (X_1, X_2) is conditionally independent of Y_1 given X_1 (see [22, Proposition 6.8]). Therefore, $E(F|X_1, X_2)$ is conditionally independent of Y_1 given X_1. It follows that (by [22, Proposition 6.6])

$$E(E(F|X_1, X_2)|X_1, Y_1) = E(E(F|X_1, X_2)|X_1)$$

Therefore,

$$\begin{aligned} E(E(F|X_1, X_2)|Y_1) &= E(E(E(F|X_1, X_2)|X_1, Y_1)|Y_1) \\ &= E(E(E(F|X_1, X_2)|X_1)|Y_1) \\ &= E(E(F|X_1)|Y_1) \end{aligned}$$

as required. □

We are now prepared to complete the proofs of Theorem 2.1 and Theorem 2.2:

Theorem 2.1. *Let I and J be arbitrary sets, let X_i be a collection of random variables indexed by $i \in I$, and let Y_j be a collection of random variables indexed by $j \in J$, and assume that these two collections together are jointly Gaussian. Then $\rho(X,Y) = r(X,Y)$.*

Proof. By Lemma 2.18, we have

$$\rho(X,Y) = \sup\{\rho(X_{I_0}, Y_{J_0}) : I_0 \subseteq I, J_0 \subseteq J, |I_0| < \infty, |J_0| < \infty\}$$

Also, by the definition of $r(X,Y)$, we have

$$r(X,Y) = \sup\{r(X_{I_0}, Y_{J_0}) : I_0 \subseteq I, J_0 \subseteq J, |I_0| < \infty, |J_0| < \infty\}$$

Therefore, to prove the theorem, it suffices to prove that $\rho(X,Y) = r(X,Y)$ in the case where X and Y are finite collections of random variables. The inequality $\rho(X,Y) \geq r(X,Y)$ is clear, so we need only to prove the reverse inequality. Applying Theorem 2.19 to the Hilbert space $V = L^2$, let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be orthonormal bases of $\mathrm{Span}(X)$ and $\mathrm{Span}(Y)$, respectively, and without loss of generality assume $m \leq n$. Let $F \in L^2_s(\sigma(X))$ and $G \in L^2_s(\sigma(Y))$ be given. For $i = 1, \ldots, m$ define

$$F_i = E(F|X_1, \ldots, X_i) - E(F|X_1, \ldots, X_{i-1})$$

Likewise, for $j = 1, \ldots, n$ define

$$G_j = E(G|Y_1, \ldots, Y_j) - E(G|Y_1, \ldots, Y_{j-1})$$

Since $E(F) = E(G) = 0$, and $F = E(F|X_1, \ldots, X_m)$, and $G = E(G|Y_1, \ldots, Y_n)$, we have

$$F = \sum_{i=1}^{m} F_i, \quad G = \sum_{j=1}^{n} G_j.$$

Given $i, j \in \{1, \ldots, m\}$ with $i > j$, we have

$$\begin{aligned}
&E(F_i|X_1, \ldots, X_j) \\
&= E(E(F|X_1, \ldots, X_i)|X_1, \ldots, X_j) - E(E(F|X_1, \ldots, X_{i-1})|X_1, \ldots, X_j) \\
&= E(F|X_1, \ldots, X_j) - E(F|X_1, \ldots, X_j) = 0
\end{aligned}$$

and therefore,

$$E(F_i F_j) = E(E(F_i F_j|X_1, \ldots, X_j))$$

$$= E(F_j E(F_i|X_1, \ldots, X_j))$$
$$= 0$$

In other words, $E(F_i F_j) = 0$ if $i \neq j$. Similarly $E(G_i G_j) = 0$ if $i \neq j$. It follows that

$$E(F^2) = E\left(\left(\sum_{i=1}^{m} F_i\right)\left(\sum_{j=1}^{m} F_j\right)\right) = \sum_{i=1}^{m}\sum_{j=1}^{m} E(F_i F_j) = \sum_{i=1}^{m} E(F_i^2)$$

and likewise $E(G^2) = \sum_{i=1}^{n} E(G_i^2)$. Since $F \in L_s(\sigma(X))$ and $G \in L_s(\sigma(Y))$, we have $E(F^2) = 1$ and $E(G^2) = 1$, so that

$$\sum_{i=1}^{m} E(F_i^2) = 1, \quad \sum_{i=1}^{n} E(G_i^2) = 1 \tag{2.12}$$

Given $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$ with $i > j$, applying Lemma 2.20 with $(X_1, \ldots, X_j)$ in place of X_1, $(X_{j+1}, \ldots, X_i)$ in place of X_2, and $(Y_1, \ldots, Y_j)$ in place Y_1, we have

$$\begin{aligned} E(F_i|Y_1, \ldots, Y_j) &= E(E(F|X_1, \ldots, X_i)|Y_1, \ldots, Y_j) - E(E(F|X_1, \ldots, X_{i-1})|Y_1, \ldots, Y_j) \\ &= E(E(F|X_1, \ldots, X_j)|Y_1, \ldots, Y_j) - E(E(F|X_1, \ldots, X_j)|Y_1, \ldots, Y_j) \\ &= 0 \end{aligned}$$

In this case,

$$\begin{aligned} E(F_i G_j) &= E(E(F_i G_j|Y_1, \ldots, Y_j)) \\ &= E(G_j E(F_i|Y_1, \ldots, Y_j)) \\ &= 0 \end{aligned}$$

Likewise, if $i < j$, we have

$$\begin{aligned} E(G_j|X_1, \ldots, X_i) &= E(E(G|Y_1, \ldots, Y_j)|X_1, \ldots, X_i) - E(E(G|Y_1, \ldots, Y_{j-1})|X_1, \ldots, X_j) \\ &= E(E(G|Y_1, \ldots, Y_j)|X_1, \ldots, X_i) - E(E(G|Y_1, \ldots, Y_j)|X_1, \ldots, X_i) \\ &= 0, \end{aligned}$$

in which case,

$$\begin{aligned} E(F_i G_j) &= E(E(F_i G_j|X_1, \ldots, X_i)) \\ &= E(F_i E(G_j|X_1, \ldots, X_i)) \\ &= 0 \end{aligned}$$

Thus $E(F_iG_j) = 0$ whenever $i \neq j$, so that

$$E(FG) = E\left(\left(\sum_{i=1}^{m} F_i\right)\left(\sum_{j=1}^{n} G_j\right)\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} E(F_iG_j) = \sum_{i=1}^{m} E(F_iG_i) \tag{2.13}$$

Now, by [22, Lemma 1.13] we may write

$$F_i = f_i(X_1, \ldots, X_i)$$
$$G_i = g_j(Y_1, \ldots, Y_i)$$

for some measurable functions f_i and g_i, for $i = 1, \ldots, m$. Given $i \in \{1, \ldots, m\}$, since $(X_1, \ldots, X_{i-1}, Y_1, \ldots, Y_{i-1})$ is independent of (X_i, Y_i), we have (by [22, Theorem 6.4])

$$\begin{aligned}
&E(F_iG_i|X_1, \ldots, X_{i-1}, Y_1, \ldots, Y_{i-1}) \\
&= E(f_i(X_1, \ldots, X_i)g_i(Y_1, \ldots, Y_i)|X_1, \ldots, X_{i-1}, Y_1, \ldots, Y_{i-1}) \\
&= E(f_i(x_1, \ldots, x_{i-1}, X_i)g_i(y_1, \ldots, y_{i-1}, Y_i))_{x_1=X_1,\ldots,x_{i-1}=X_{i-1},y_1=Y_1,\ldots,y_{i-1}=Y_{i-1}}
\end{aligned} \tag{2.14}$$

Here, since X_i is independent of $(X_1, \ldots, X_{i-1})$ we have (again by [22, Theorem 6.4])

$$\begin{aligned}
&E(f_i(x_1, \ldots, x_{i-1}, X_i))_{x_1=X_1,\ldots,x_{i-1}=X_{i-1}} \\
&= E(f_i(X_1, \ldots, X_i)|X_1, \ldots, X_{i-1}) \\
&= E(F_i|X_1, \ldots, X_{i-1}) \\
&= E(F|X_1, \ldots, X_{i-1}) - E(F|X_1, \ldots, X_{i-1}) \\
&= 0,
\end{aligned}$$

and similarly,

$$E(g_i(y_1, \ldots, y_{i-1}, Y_i))_{y_1=Y_1,\ldots,y_{i-1}=Y_{i-1}} = 0.$$

Equation (2.14) then becomes

$$\begin{aligned}
&E(F_iG_i|X_1, \ldots, X_{i-1}, Y_1, \ldots, Y_{i-1}) \\
&= \mathrm{Cov}(f_i(x_1, \ldots, x_{i-1}, X_i), g_i(y_1, \ldots, y_{i-1}, Y_i))_{x_1=X_1,\ldots,x_{i-1}=X_{i-1},y_1=Y_1,\ldots,y_{i-1}=Y_{i-1}}
\end{aligned} \tag{2.15}$$

By Theorem 2.15, we have $\rho(X_i, Y_i) = r(X_i, Y_i) \leq r(X, Y)$, so that

$$\begin{aligned}
&|\,\mathrm{Cov}(f_i(x_1, \ldots, x_{i-1}, X_i), g_i(y_1, \ldots, y_{i-1}, Y_i))| \\
&\leq r(X, Y)\sqrt{\mathrm{Var}(f_i(x_1, \ldots, x_{i-1}, X_i))}\sqrt{\mathrm{Var}(g_i(y_1, \ldots, y_{i-1}, Y_i))} \\
&= r(X, Y)\sqrt{E[(f_i(x_1, \ldots, x_{i-1}, X_i))^2]}\sqrt{E[(g_i(y_1, \ldots, y_{i-1}, Y_i))^2]}.
\end{aligned} \tag{2.16}$$

We have (again by [22, Theorem 6.4])

$$E[(f_i(x_1,\dots,x_{i-1},X_i))^2]_{x_1=X_1,\dots,x_{i-1}=X_{i-1}} = E[(f_i(X_1,\dots,X_{i-1},X_i))^2|X_1,\dots,X_{i-1}] \\ = E[F_i^2|X_1,\dots,X_{i-1}], \tag{2.17}$$

and likewise,

$$E[(g_i(y_1,\dots,y_{i-1},Y_i))^2]_{y_1=Y_1,\dots,y_{i-1}=Y_{i-1}} = E[G_i^2|Y_1,\dots,Y_{i-1}], \tag{2.18}$$

From (2.15), (2.16), (2.17), and (2.18) we then deduce

$$|E(F_iG_i|X_1,\dots,X_{i-1},Y_1,\dots,Y_{i-1})| \\ \le r(X,Y)\sqrt{E\left[F_i^2|X_1,\dots,X_{i-1}\right]}\sqrt{E\left[G_i^2|Y_1,\dots,Y_{i-1}\right]}$$

and hence, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
|E(F_iG_i)| &= |E(E(F_iG_i|X_1,\dots,X_{i-1},Y_1,\dots,Y_{i-1}))| \\
&\le E|E(F_iG_i|X_1,\dots,X_{i-1},Y_1,\dots,Y_{i-1})| \\
&\le E\left(r(X,Y)\sqrt{E\left[F_i^2|X_1,\dots,X_{i-1}\right]}\sqrt{E\left[G_i^2|Y_1,\dots,Y_{i-1}\right]}\right) \\
&= r(X,Y)E\left(\sqrt{E\left[F_i^2|X_1,\dots,X_{i-1}\right]}\sqrt{E\left[G_i^2|Y_1,\dots,Y_{i-1}\right]}\right) \\
&\le r(X,Y)\sqrt{E\left(E\left[F_i^2|X_1,\dots,X_{i-1}\right]\right)E\left(E\left[G_i^2|Y_1,\dots,Y_{i-1}\right]\right)} \\
&= r(X,Y)\sqrt{E\left(F_i^2\right)E\left(G_i^2\right)}
\end{aligned}$$

Applying this bound to (2.13), using the Cauchy-Schwarz inequality once again, and applying (2.12), we obtain

$$\begin{aligned}
|E(FG)| &= \left|\sum_{i=1}^m E(F_iG_i)\right| \\
&\le \sum_{i=1}^m |E(F_iG_i)| \\
&\le \sum_{i=1}^m r(X,Y)\sqrt{E\left(F_i^2\right)E\left(G_i^2\right)} \\
&= r(X,Y)\sum_{i=1}^m \sqrt{E\left(F_i^2\right)}\sqrt{E\left(G_i^2\right)} \\
&\le r(X,Y)\sqrt{\left(\sum_{i=1}^m E(F_i^2)\right)\left(\sum_{i=1}^m E(G_i^2)\right)}
\end{aligned}$$

$$\leq r(X,Y)\sqrt{\left(\sum_{i=1}^{m} E(F_i^2)\right)\left(\sum_{i=1}^{n} E(G_i^2)\right)}$$
$$= r(X,Y)$$

Taking the supremum over all $F \in L_s^2(\sigma(X))$ and $G \in L_s^2(\sigma(Y))$, we obtain $\rho(X,Y) \geq r(X,Y)$, as required. □

Theorem 2.2. *Let I and J be arbitrary sets, let X_i be a collection of random variables indexed by $i \in I$, and let Y_j be a collection of random variables indexed by $j \in J$, and assume that these two collections together are jointly Gaussian. Then $4\alpha(X,Y) \leq \rho(X,Y) \leq 2\pi\alpha(X,Y)$*

Proof. The first inequality was already given in 2.2. Let $\epsilon > 0$ be given. By the definition of $r(X,Y)$, we may choose linear combinations $Z = \sum_i a_i X_i$ and $W = \sum_j b_j Y_j$ such that $\mathrm{Corr}(Z,W) \geq r(X,Y) - \epsilon$. Without loss of generality, assume $\mathrm{Var}(Z) = 1 = \mathrm{Var}(W)$. Writing $r = \mathrm{Corr}(Z,W)$, we have

$$Z = N_1$$
$$W = rN_1 + \sqrt{1-r^2}N_2$$

for independent standard Gaussian random variables N_1, N_2. Now, write (N_1, N_2) in polar form:

$$N_1 = R\cos\Theta$$
$$N_2 = R\sin\Theta$$

with $R \geq 0$ and $-\pi \leq \Theta < \pi$; this defines R and Θ uniquely as random variables, except on the null event $(N_1, N_2) = (0,0)$, where Θ may be defined arbitrarily, so define $\Theta = 0$ in that case. The rotational symmetry of the distribution of (N_1, N_2) implies that Θ has a uniform distribution on $[-\pi, \pi]$. Now, define events $A = \{Z \geq 0\}$ and $B = \{W \geq 0\}$.

$$\begin{aligned} P(A \cap B) &= P\left(N_1 \geq 0 \cap \tan\Theta \geq \frac{-r}{\sqrt{1-r^2}}\right) \\ &= P\left(\tan^{-1}\left(\frac{-r}{\sqrt{1-r^2}}\right) \leq \Theta \leq \frac{\pi}{2}\right) \\ &= \frac{1}{2\pi}\left(\frac{\pi}{2} - \tan^{-1}\left(\frac{-r}{\sqrt{1-r^2}}\right)\right) \\ &= \frac{1}{4} + \frac{1}{2\pi}\tan^{-1}\left(\frac{r}{\sqrt{1-r^2}}\right) \end{aligned}$$

$$= \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(r)$$
$$= P(A)P(B) + \frac{1}{2\pi}\sin^{-1}(r)$$

Therefore, we have

$$\frac{1}{2\pi}\sin^{-1}(r) = P(A \cap B) - P(A)P(B) \leq \alpha(X, Y)$$

and hence

$$\rho(X, Y) - \epsilon = r(X, Y) - \epsilon \leq r \leq \sin^{-1}(r) \leq 2\pi\alpha(X, Y)$$

Taking the limit as $\epsilon \to 0$, we obtain $\rho(X, Y) \leq 2\pi\alpha(X, Y)$. □

The following theorem was shown by Bradley [4, Theorem 28.19]:

Theorem 2.3. *Let $Z(\boldsymbol{s})$ be a second-order stationary random field on $\mathbb{Z}^n$, and assume that it has a continuous, positive spectral density $f(\boldsymbol{\omega})$. Then Z is r^*-mixing.*

Proof. We identify $\mathbb{T}^n$ with the dual space of $\mathbb{Z}^n$, by associating $\boldsymbol{\omega} \in \mathbb{T}^n$ with the character $\chi(\boldsymbol{x}) = e^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{x}}$. Since f is continuous on the compact set $\mathbb{T}^n$, it has a minimum value $M > 0$. Let A be the set of functions $g : \mathbb{T}^n \to \mathbb{R}$ of the form

$$g(\boldsymbol{\omega}) = \sum_{\boldsymbol{x}\in\mathbb{Z}^n} c_{\boldsymbol{x}} e^{\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{x}}$$

where only finitely many coefficients $c_{\boldsymbol{x}} \in \mathbb{R}$ are nonzero, and where $c_{-\boldsymbol{x}} = c_{\boldsymbol{x}}$ for all $\boldsymbol{x}$. Then A is an algebra which separates points and contains all constant functions, so by the Stone-Weierstrass theorem, A is dense in the space of all real-valued continuous functions on $\mathbb{T}^n$, with respect to the supremum norm. In particular, given $\epsilon > 0$, the spectral density f may be uniformly approximated by a function in A, so there exists a function $g \in A$ such that

$$|g(\boldsymbol{\omega}) - f(\boldsymbol{\omega})| \leq \epsilon M$$

for all $\boldsymbol{\omega}$. Define $C_1 : \mathbb{Z}^n \to \mathbb{R}$ by $\boldsymbol{x} \mapsto c_{-\boldsymbol{x}}$. Then C_1 has compact support, and we have $g(\boldsymbol{\omega}) = \hat{C}_1$, hence

$$|\hat{C}_1(\boldsymbol{\omega}) - f(\boldsymbol{\omega})| = |g(\boldsymbol{\omega}) - f(\boldsymbol{\omega})| \leq \epsilon M \leq \epsilon f(\boldsymbol{\omega})$$

for all $\boldsymbol{\omega} \in \mathbb{T}^n$. Therefore, by Theorem 2.8, $r^*(K) \leq \epsilon$, where K is the support of C_1. In particular, if we let $h = \max_{\boldsymbol{x}\in K} \|\boldsymbol{x}\|$, then $r^*(h) \leq \epsilon$. Since ϵ was arbitrary, this proves that Z is r^* mixing. □

Theorem 2.5. *Let $Z(\boldsymbol{s})$ be a second-order stationary random field over $\mathbb{R}^n$ with covariance function $C(\boldsymbol{h})$ and spectral density $f(\boldsymbol{\omega})$, and let $R > 0$ be given. If there exists a bounded function $C_1(\boldsymbol{h})$ with support contained in B_R such that the Fourier transform $\hat{C}_1$ satisfies*

$$|\hat{C}_1(\boldsymbol{\omega}) - f(\boldsymbol{\omega})| \leq \epsilon f(\boldsymbol{\omega}) \tag{2.19}$$

then $r^*(R) \leq \epsilon$.

Proof. Let A, B be finite subsets of $\mathbb{R}^n$ with $d(A, B) \geq R$, and let $c_{\boldsymbol{x}}$ for $\boldsymbol{x} \in A$ and $c_{\boldsymbol{y}}$ for $\boldsymbol{y} \in B$ be given real numbers. Define random variables

$$X = \sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} Z(\boldsymbol{x}), \quad Y = \sum_{\boldsymbol{y} \in B} c_{\boldsymbol{y}} Z(\boldsymbol{y}).$$

Assuming without loss of generality that $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, we must show $|\operatorname{Cov}(X, Y)| \leq \epsilon$. In fact we have

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \operatorname{Cov}\left(\sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} Z(\boldsymbol{x}), \sum_{\boldsymbol{y} \in B} c_{\boldsymbol{y}} Z(\boldsymbol{x}) \right) \\
&= \sum_{\boldsymbol{x} \in A, \boldsymbol{y} \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \operatorname{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) \\
&= \sum_{\boldsymbol{x} \in A, \boldsymbol{y} \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} C(\boldsymbol{y} - \boldsymbol{x}) \\
&= \sum_{\boldsymbol{x} \in A, \boldsymbol{y} \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\boldsymbol{\omega}} f(\boldsymbol{\omega}) \, d\boldsymbol{\omega} \\
&= \sum_{\boldsymbol{x} \in A, \boldsymbol{y} \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \left(\int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\boldsymbol{\omega}} \hat{C}_1(\boldsymbol{\omega}) \, d\boldsymbol{\omega} + \int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\omega} (f(\boldsymbol{\omega}) - \hat{C}_1(\boldsymbol{\omega})) \, d\boldsymbol{\omega} \right)
\end{aligned} \tag{2.20}$$

Now the assumptions imply that both C_1 and $\hat{C}_1$ are in $L^1(\mathbb{R}^n)$, so we may apply Fourier inversion to obtain that for all $\boldsymbol{x} \in A$ and $\boldsymbol{y} \in B$,

$$\int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\omega} \hat{C}_1(\omega) \, d\omega = C_1(\boldsymbol{y} - \boldsymbol{x}) = 0$$

since $\|\boldsymbol{y} - \boldsymbol{x}\| \geq R$ implies that $\boldsymbol{x} - \boldsymbol{y}$ is outside the support of C_1. Applying this to (2.20), we have

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \sum_{x \in A, y \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\boldsymbol{\omega}} (f(\boldsymbol{\omega}) - \hat{C}_1(\boldsymbol{\omega})) \, d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^n} \left(\sum_{\boldsymbol{x} \in A, \boldsymbol{y} \in B} c_{\boldsymbol{x}} c_{\boldsymbol{y}} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\boldsymbol{\omega}} \right) (f(\boldsymbol{\omega}) - \hat{C}_1(\boldsymbol{\omega})) \, d\boldsymbol{\omega}
\end{aligned}$$

$$= \int_{\mathbb{R}^n} \left(\sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} e^{-\boldsymbol{x} \cdot \boldsymbol{\omega} \mathrm{i}} \right) \left(\sum_{\boldsymbol{y} \in B} c_{\boldsymbol{y}} e^{\boldsymbol{y} \cdot \boldsymbol{\omega} \mathrm{i}} \right) (f(\boldsymbol{\omega}) - \hat{C}_1(\boldsymbol{\omega}))\ d\boldsymbol{\omega}$$

By (2.19), this implies

$$|\operatorname{Cov}(X, Y)| \leq \epsilon \int_{\mathbb{R}^n} \left| \sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} e^{-\boldsymbol{x} \cdot \boldsymbol{\omega} \mathrm{i}} \right| \left| \sum_{\boldsymbol{y} \in B} c_{\boldsymbol{y}} e^{\boldsymbol{y} \cdot \boldsymbol{\omega} \mathrm{i}} \right| f(\boldsymbol{\omega})\ d\boldsymbol{\omega} \tag{2.21}$$

On the other hand, we have

$$\begin{aligned}
1 = \operatorname{Var}(X) &= \operatorname{Cov}\left(\sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} Z(\boldsymbol{x}), \sum_{\boldsymbol{y} \in A} c_{\boldsymbol{y}} Z(\boldsymbol{y}) \right) \\
&= \sum_{\boldsymbol{x}, \boldsymbol{y} \in A} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \operatorname{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{y})) \\
&= \sum_{\boldsymbol{x}, \boldsymbol{y} \in A} c_{\boldsymbol{x}} c_{\boldsymbol{y}} C(\boldsymbol{y} - \boldsymbol{x}) \\
&= \sum_{\boldsymbol{x}, \boldsymbol{y} \in A} c_{\boldsymbol{x}} c_{\boldsymbol{y}} \int_{\mathbb{R}^n} e^{(\boldsymbol{y} - \boldsymbol{x}) \cdot \mathrm{i}\boldsymbol{\omega}} f(\boldsymbol{\omega})\ d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^n} \left(\sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} e^{-\boldsymbol{x} \cdot \boldsymbol{\omega} \mathrm{i}} \right) \left(\sum_{\boldsymbol{y} \in A} c_{\boldsymbol{y}} e^{\boldsymbol{y} \cdot \boldsymbol{\omega} \mathrm{i}} \right) f(\boldsymbol{\omega})\ d\boldsymbol{\omega} \\
&= \int_{\mathbb{R}^n} \left| \sum_{\boldsymbol{x} \in A} c_{\boldsymbol{x}} e^{-\boldsymbol{x} \cdot \boldsymbol{\omega} \mathrm{i}} \right|^2 f(\boldsymbol{\omega})\ d\boldsymbol{\omega}
\end{aligned}$$

and similarly $\int_{\mathbb{R}^n} \left| \sum_{\boldsymbol{y} \in B} c_{\boldsymbol{y}} e^{\boldsymbol{y} \cdot \boldsymbol{\omega} \mathrm{i}} \right|^2 f(\boldsymbol{\omega})\ d\boldsymbol{\omega} = 1$, so applying the Cauchy-Schwarz inequality to (2.21), we conclude that $|\operatorname{Cov}(X, Y)| \leq \epsilon$, as desired. □

Lemma 2.21. *Suppose that $g \in L^1(\mathbb{R}^2)$ is circularly symmetric, i.e., $g(\boldsymbol{x}) = g_0(\|\boldsymbol{x}\|)$ for some function g_0. Then the Fourier transform of g may be expressed*

$$\hat{g}(\boldsymbol{\omega}) = \frac{1}{2\pi} \int_0^\infty J_0(\|\boldsymbol{\omega}\| r) r g_0(r)\ dr \tag{2.22}$$

where J_0 is the zeroth-order Bessel function of the first kind:

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{-\mathrm{i}x \sin\theta}\ d\theta$$

Proof. First observe that $\hat{g}$ is circularly symmetric: for given $\boldsymbol{\omega}, \boldsymbol{\omega}'$ with $\|\boldsymbol{\omega}\| = \|\boldsymbol{\omega}'\|$, there is an orthogonal matrix A such that $\boldsymbol{\omega}' = A\boldsymbol{\omega}$, and applying a change of variables $\boldsymbol{x} = A\boldsymbol{y}$ to the definition of the Fourier transform, we obtain

$$\hat{g}(A\boldsymbol{\omega}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-\mathrm{i}A\boldsymbol{\omega} \cdot \boldsymbol{x}} g(\boldsymbol{x})\ d\boldsymbol{x}$$

$$= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-\mathrm{i}A\boldsymbol{\omega}\cdot A\boldsymbol{y}} g(A\boldsymbol{y})\ d\boldsymbol{y}$$
$$= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{y}} g(\boldsymbol{y})\ d\boldsymbol{y} = \hat{g}(\boldsymbol{\omega})$$

Thus it suffices to prove (2.22) in the case where $\boldsymbol{\omega} = (0, \|\boldsymbol{\omega}\|)$, and for this we use polar coordinates to compute

$$\begin{aligned}
\hat{g}(0, \|\boldsymbol{\omega}\|) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{x}} g(\boldsymbol{x})\ d\boldsymbol{x} \\
&= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} e^{-\mathrm{i}\|\boldsymbol{\omega}\|x_2} g_0(\|\boldsymbol{x}\|)\ d\boldsymbol{x} \\
&= \frac{1}{(2\pi)^2} \int_0^\infty \int_0^{2\pi} e^{-\mathrm{i}\|\boldsymbol{\omega}\| r\sin\theta} g_0(r) r\ d\theta\ dr \\
&= \frac{1}{(2\pi} \int_0^\infty \left(\frac{1}{2\pi} \int_0^{2\pi} e^{-\mathrm{i}\|\boldsymbol{\omega}\| r\sin\theta}\ d\theta \right) g_0(r) r\ dr \\
&= \frac{1}{\pi} \int_0^\infty J_0(\|\boldsymbol{\omega}\| r) r g_0(r)\ d\theta\ dr
\end{aligned}$$

as desired. □

Theorem 2.7. *Let $Z(\boldsymbol{s})$ be a random field over $\mathbb{R}^2$ with isotropic covariance function $C(h) = \sigma^2 e^{-h/h_0}$. Then Z is r^*-mixing. More precisely, the r^* mixing coefficients satisfy the bound*

$$r^*(R) \le e^{-R/h_0} \left(2 + 140\pi^2\sqrt{5} \left(\frac{R}{h_0} \right)^2 + \frac{\pi^2}{2} + \frac{\pi^2\sqrt{10}}{3} \left(\frac{5R}{h_0} + 20 \right) \right)$$

Proof. Without loss of generality, we may assume $\sigma^2 = 1$ and $h_0 = 1$. We will apply Theorem 2.5 with $C_1(\boldsymbol{h}) = e^{-\|\boldsymbol{h}\|}(1 - e^{\|\boldsymbol{h}\|-R})^2$, for $0 \le \|\boldsymbol{h}\| \le R$, and $C_1(\boldsymbol{h}) = 0$ for $\|\boldsymbol{h}\| > R$. We compute the isotropic spectral density, using Lemma 2.21,

$$\begin{aligned}
f(\omega) &= \frac{1}{2\pi} \int_0^\infty J_0(\omega r) r C(r)\ dr \\
&= \frac{1}{(2\pi)^2} \int_0^\infty \int_0^{2\pi} e^{-\mathrm{i}wr\sin\theta} r e^{-r}\ d\theta\ dr \\
&= \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^\infty r e^{-r(1+\mathrm{i}w\sin\theta)}\ dr\ d\theta \\
&= \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{1}{(1+\mathrm{i}w\sin\theta)^2}\ d\theta
\end{aligned}$$

To carry out the computation, we use the substitution $z = \omega e^{i\theta}$, giving a contour integral along the circle $|z| = \omega$, oriented clockwise:

$$f(\omega) = \frac{1}{(2\pi)^2} \oint \frac{1}{\mathrm{i}z(1 + \frac{z-\bar{z}}{2})^2}$$

$$= \frac{1}{(2\pi)^2 \mathrm{i}} \oint \frac{z}{(\frac{1}{2}z^2 + z - \frac{\omega^2}{2})^2}$$
$$= \frac{1}{\pi^2 \mathrm{i}} \oint \frac{z}{(z - z_1)^2 (z - z_2)^2}$$

where $z_1 = -1 + \sqrt{1 + \omega^2}$ and $z_2 = -1 - \sqrt{1 + \omega^2}$. Then $|z_1| < \omega$ and $|z_2| > \omega$, so z_1 is enclosed by the contour while z_2 is not. We can then complete the computation using the residue theorem,

$$\begin{aligned} f(\omega) &= \frac{2}{\pi} \operatorname{Res}_{z=z_1} \frac{z}{(z - z_1)^2 (z - z_2)^2} \\ &= \frac{2}{\pi} \frac{d}{dz}\bigg|_{z=z_1} \frac{z}{(z - z_2)^2} \\ &= \frac{2}{\pi} \left(\frac{1}{(z_1 - z_2)^2} - \frac{2z_1}{(z_1 - z_2)^3} \right) \\ &= \frac{-2(z_1 + z_2)}{\pi (z_1 - z_2)^3} \\ &= \frac{4}{\pi (2\sqrt{1 + \omega^2})^3} \\ &= \frac{1}{2\pi (1 + \omega^2)^{3/2}} \end{aligned} \tag{2.23}$$

For reference below, we also note another integral form for $f(\omega)$, obtained by taking the real part of the integrand and substituting $u = \omega \sin\theta$:

$$\begin{aligned} f(\omega) &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{1}{(1 + \mathrm{i}w \sin\theta)^2} \, d\theta \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{(1 - \mathrm{i}w \sin\theta)^2}{(1 + w^2 \sin^2\theta)^2} \, d\theta \\ &= \frac{1}{(2\pi)^2} \int_0^{2\pi} \frac{1 - w^2 \sin^2\theta}{(1 + w^2 \sin^2\theta)^2} \, d\theta \\ &= \frac{1}{\pi^2} \int_0^{\pi/2} \frac{1 - w^2 \sin^2\theta}{(1 + w^2 \sin^2\theta)^2} \, d\theta \\ &= \frac{1}{\pi^2} \int_0^{\omega} \frac{1 - u^2}{(1 + u^2)^2 \sqrt{\omega^2 - u^2}} \, du \end{aligned} \tag{2.24}$$

We now turn to the calculation of $\hat{C}_1(\omega)$:

$$\begin{aligned} \hat{C}_1(\omega) &= \frac{1}{2\pi} \int_0^{\infty} J_0(\omega r) r C_1(r) \, dr \\ &= \frac{1}{(2\pi)^2} \int_0^{R} \int_0^{2\pi} e^{-\mathrm{i}\omega r \sin\theta} r e^{-r} (1 - e^{r-R})^2 \, d\theta \, dr \\ &= \frac{1}{\pi^2} \int_0^{R} \int_0^{\pi/2} \cos(\omega r \sin\theta) r e^{-r} (1 - e^{r-R})^2 \, d\theta \, dr \end{aligned}$$

We apply the substitution $u = \omega \sin\theta$ and then use symbolic algebra software to evaluate the integral:

$$\begin{aligned}\hat{C}_1(\omega) &= \frac{1}{\pi^2}\int_0^R \int_0^1 \frac{\cos(\omega r u) r e^{-r}(1-e^{r-R})^2}{\sqrt{1-u^2}}\, du\, dr \\ &= \frac{1}{\pi^2}\int_0^\omega \frac{1}{\sqrt{\omega^2-u^2}} \int_0^R \cos(ru) r e^{-r}(1-e^{r-R})^2\, dr\, du \\ &= \frac{1}{\pi^2}\int_0^\omega \left[\frac{1-u^2}{(1+u^2)^2\sqrt{\omega^2-u^2}}(1-e^{-R})^2 - 2e^{-R}\left(I_1(u)+I_2(u)\right)\right] du,\end{aligned}$$

where

$$\begin{aligned}I_1(u) &= \frac{R\sin(Ru)}{u(1+u^2)\sqrt{\omega^2-u^2}}, \\ I_2(u) &= \frac{(3u^2+1)(\cos(Ru)-1)}{u^2(1+u^2)^2\sqrt{\omega^2-u^2}}.\end{aligned}$$

Applying (2.24), this becomes

$$\hat{C}_1(\omega) = f(\omega)(1-e^{-R})^2 - 2e^{-R}\left(\int_0^\omega I_1(u)\, du + \int_0^\omega I_2(u)\, du\right) \tag{2.25}$$

We now need to estimate the two integrals in (2.25). We may express them in terms of Cauchy principal value integrals on $[-\omega, \omega]$:

$$\int_0^\omega I_1(u)\, du = \frac{1}{2\mathrm{i}}\int_{-\omega}^\omega I_1^*(u)\, du, \tag{2.26}$$

$$\int_0^\omega I_2(u)\, du = \frac{1}{2}\int_{-\omega}^\omega I_2^*(u)\, du, \tag{2.27}$$

where

$$\begin{aligned}I_1^*(u) &= \frac{Re^{\mathrm{i}Ru}}{u(1+u^2)\sqrt{\omega^2-u^2}}, \\ I_2^*(u) &= \frac{(3u^2+1)(e^{\mathrm{i}Ru}-1)}{u^2(1+u^2)^2\sqrt{\omega^2-u^2}}.\end{aligned}$$

For the moment, assume that $\omega > 1$. We will consider the closed contour γ consisting of the line segment γ_1 from $-\omega$ to ω followed by the semicircular arc γ_2, centered at 0, in the upper half plane from ω to $-\omega$, as shown in Figure 2.1.

Next consider the modified contour γ_ϵ^* obtained from γ by indenting the contour by a semicircle of radius ϵ at the origin and by circular arcs of radius ϵ at $-\omega$ and ω, as shown in Figure 2.2. Taking the standard branch of $\sqrt{z}$ cut along the negative real axis, both I_1^* and I_2^* are analytic on and inside the contour γ_ϵ^* except for the pole at $z = \mathrm{i}$, so we can apply the residue theorem:

$$\oint_{\gamma_\epsilon^*} I_j^*(z)\, dz = 2\pi\mathrm{i}\,\mathrm{Res}_{z=\mathrm{i}}\, I_j^*(z) \tag{2.28}$$

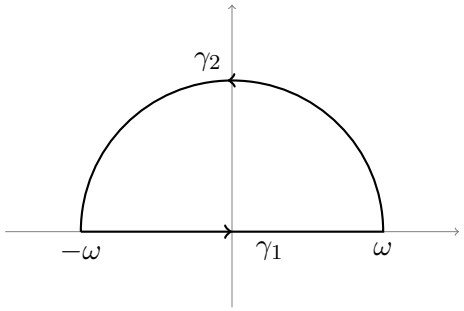

Figure 2.1: The closed contour γ

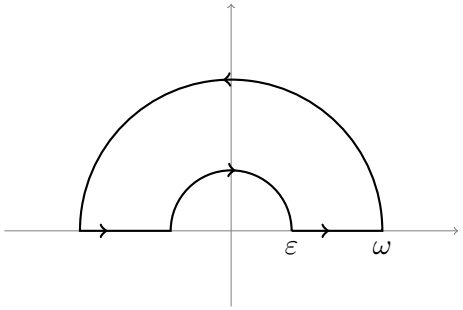

Figure 2.2: The closed contour γ_ϵ^*

On the other hand, we have the decomposition

$$\oint_{\gamma_\epsilon^*} I_j^*(z)\ dz = \sum_{k=1}^{6} \int_{\gamma_{\epsilon,k}^*} I_j^*(z)\ dz \tag{2.29}$$

As $I_1^*(z)$ and $I_2^*(z)$ each have only a simple pole $z = 0$, we have

$$\lim_{\epsilon\to 0} \int_{\gamma_{\epsilon,3}^*} I_j^*(z)\ dz = -\pi \mathrm{i}\, \mathrm{Res}_{z=0}\, I_j^*(z) \tag{2.30}$$

Also

$$\lim_{\epsilon\to 0} \int_{\gamma_{\epsilon,1}^*} I_j^*(z)\ dz = \lim_{\epsilon\to 0} \int_{\gamma_{\epsilon,5}^*} I_j^*(z)\ dz = 0 \tag{2.31}$$

$$\lim_{\epsilon\to 0} \left(\int_{\gamma_{\epsilon,2}^*} I_j^*(z)\ dz + \int_{\gamma_{\epsilon,4}} I_j^*(z)\ dz \right) = \int_{\gamma_1} I_j^*(z)\ dz \tag{2.32}$$

$$\lim_{\epsilon\to 0} \int_{\gamma_{\epsilon,6}^*} I_j^*(z)\ dz = \int_{\gamma_2} I_j^*(z)\ dz \tag{2.33}$$

where $\int_{\gamma_1} I_j^*(z)\ dz = \int_{-\omega}^{\omega} I_j^*(z)\ dz$ is to be interpreted as the Cauchy principal value integral, as before. Taking the limit as $\epsilon \to 0$ in (2.28) and substituting equations (2.29) through (2.33), we obtain

$$\int_{-\omega}^{\omega} I_j^*(u)\ du = 2\pi \mathrm{i}\, \mathrm{Res}_{z=i}\, I_j^*(z) + \pi \mathrm{i}\, \mathrm{Res}_{z=0}\, I_j^*(z) - \int_{\gamma_2} I_j^*(z)\ dz \tag{2.34}$$

Here we calculate the residues

$$\mathrm{Res}_{z=\mathrm{i}}\, I_1^*(z) = \left.\frac{Re^{\mathrm{i}Rz}}{z(z+\mathrm{i})\sqrt{\omega^2-z^2}}\right|_{z=\mathrm{i}} = \frac{-Re^{-R}}{2\sqrt{\omega^2+1}}$$

$$\mathrm{Res}_{z=0}\, I_1^*(z) = \left.\frac{Re^{\mathrm{i}Rz}}{(1+z^2)\sqrt{\omega^2-z^2}}\right|_{z=0} = \frac{R}{\omega}$$

$$\begin{aligned}\mathrm{Res}_{z=\mathrm{i}}\, I_2^*(z) &= \left.\frac{d}{dz}\right|_{z=\mathrm{i}} \frac{(3z^2+1)(e^{\mathrm{i}Rz}-1)}{z^2(z+\mathrm{i})^2\sqrt{\omega^2-z^2}} \\ &= \frac{-\mathrm{i}Re^{-R}}{2\sqrt{\omega^2+1}} - \frac{\mathrm{i}(e^{-R}-1)}{2(\omega^2+1)^{3/2}} \\ &= \frac{-\mathrm{i}Re^{-R}}{2\sqrt{\omega^2+1}} - \frac{\mathrm{i}\pi}{2}(e^{-R}-1)f(\omega)\end{aligned}$$

$$\begin{aligned}\mathrm{Res}_{z=0}\, I_2^*(z) &= \left.\frac{d}{dz}\right|_{z=0} \frac{(3z^2+1)(e^{\mathrm{i}Rz}-1)}{(1+z^2)^2\sqrt{\omega^2-z^2}} \\ &= \left.\frac{d}{dz}\right|_{z=0} \frac{(3z^2+1)(e^{\mathrm{i}Rz}-1)}{(1+z^2)^2\sqrt{\omega^2-z^2}} \\ &= \frac{\mathrm{i}R}{\omega}\end{aligned}$$

Substituting into (2.26), (2.27), and (2.34), we obtain

$$\begin{aligned}\int_0^\omega I_1(u)\, du &= \frac{1}{2\mathrm{i}}\left(\frac{-\pi\mathrm{i}Re^{-R}}{\sqrt{\omega^2+1}} + \frac{\pi\mathrm{i}R}{\omega} - \int_{\gamma_2} I_1^*(z)\, dz\right) \\ \int_0^\omega I_2(u)\, du &= \frac{1}{2}\left(\frac{\pi Re^{-R}}{\sqrt{\omega^2+1}} - \frac{\pi^2}{2}(e^{-R}-1)f(\omega) - \frac{\pi R}{\omega} - \int_{\gamma_2} I_2^*(z)\, dz\right)\end{aligned}$$

Substituting into (2.25), we obtain

$$\begin{aligned}\hat{C}_1(\omega) &= f(\omega)(1-e^{-R})^2 - 2e^{-R}\left(\int_0^\omega I_1(u)\, du + \int_0^\omega I_2(u)\, du\right) \\ &= f(\omega)(1-e^{-R})^2 + e^{-R}\left(\frac{\pi^2}{2}(e^{-R}-1)f(\omega) - \mathrm{i}\int_{\gamma_2} I_1^*(z)\, dz + \int_{\gamma_2} I_2^*(z)\, dz\right) \quad (2.35)\end{aligned}$$

Since $|e^{iRz}| \le 1$ in the upper half plane, we may bound the semicircular contour integrals as follows,

$$\begin{aligned}\left|\int_{\gamma_2} I_1^*(z)\, dz\right| &= \left|\int_{\gamma_2} \frac{Re^{iRz}}{z(1+z^2)\sqrt{\omega^2-z^2}}\, dz\right| \\ &\le \int_0^\pi \frac{R\omega\, d\theta}{\omega(\omega^2-1)\sqrt{|\omega^2-(\omega e^{\pi i\theta})^2|}} \\ &\le \frac{R}{\omega(\omega^2-1)}\int_0^\pi \frac{d\theta}{\sqrt{2\sin(\theta)}}\end{aligned}$$

$$
\begin{aligned}
&= \frac{2R}{\omega(\omega^2-1)} \int_0^{\pi/2} \frac{d\theta}{\sqrt{2\sin(\theta)}} \\
&\leq \frac{2R}{\omega(\omega^2-1)} \int_0^{\pi/2} \frac{d\theta}{\sqrt{2(\frac{2}{\pi}\theta)}} \\
&= \frac{R\sqrt{\pi}}{\omega(\omega^2-1)} \int_0^{\pi/2} \frac{d\theta}{\sqrt{\theta}} \\
&= \frac{\pi R\sqrt{2}}{\omega(\omega^2-1)}
\end{aligned}
$$

$$
\begin{aligned}
\left| \int_{\gamma_2} I_2^*(z)\, dz \right| &= \left| \int_{\gamma_2} \frac{(3z^2+1)(e^{iRz}-1)}{z^2(1+z^2)^2\sqrt{\omega^2-z^2}}\, dz \right| \\
&\leq \int_0^{\pi} \frac{2(3\omega^2-1)}{\omega^2(\omega^2-1)^2\sqrt{|\omega^2-(\omega e^{\pi i\theta})^2|}}\, d\theta \\
&\leq \frac{2\pi\sqrt{2}(3\omega^2-1)}{\omega^2(\omega^2-1)^2}
\end{aligned}
$$

By (2.23), for $\omega \geq 2$, these bounds imply

$$
\begin{aligned}
\left| \frac{1}{f(\omega)} \int_{\gamma_2} I_1^*(z)\, dz \right| &\leq \frac{2\pi^2 R\sqrt{2}(\omega^2+1)^{3/2}}{\omega(\omega^2-1)} \\
&\leq \frac{2\pi^2 R\sqrt{2}(1+\frac{1}{\omega^2})^{3/2}}{1-\frac{1}{\omega^2}} \\
&\leq \frac{5\pi^2 R\sqrt{10}}{3}
\end{aligned}
$$

$$
\begin{aligned}
\left| \frac{1}{f(\omega)} \int_{\gamma_2} I_2^*(z)\, dz \right| &\leq \frac{4\pi^2\sqrt{2}(3\omega^2-1)(\omega^2+1)^{3/2}}{\omega^2(\omega^2-1)^2} \\
&= \frac{4\pi^2\sqrt{2}(3-\frac{1}{\omega^2})(1+\frac{1}{\omega^2})^{3/2}}{\omega(1-\frac{1}{\omega^2})^2} \\
&\leq \frac{20\pi^2\sqrt{10}}{3}
\end{aligned}
$$

Applying these bounds to (2.35), we conclude that for $\omega \geq 2$,

$$
\begin{aligned}
\left| 1 - \frac{\hat{C}_1(\omega)}{f(\omega)} \right| &\leq 1 - (1-e^{-R})^2 + e^{-R}\left(\frac{\pi^2}{2}(1-e^{-R}) + \frac{\pi^2\sqrt{10}}{3}(5R+20) \right) \\
&\leq e^{-R}\left(2 + \frac{\pi^2}{2} + \frac{\pi^2\sqrt{10}}{3}(5R+20) \right) \qquad (2.36)
\end{aligned}
$$

On the other hand, for $\omega \leq 2$,

$$
\left| \int_0^{\omega} I_1(u)\, du \right| = \left| \int_0^{\omega} \frac{R\sin(Ru)}{u(1+u^2)\sqrt{\omega^2-u^2}}\, du \right|
$$

$$\begin{aligned}
&\leq \int_0^\omega \frac{R^2}{(1+u^2)\sqrt{\omega^2-u^2}}\ du \\
&\leq R^2 \int_0^\omega \frac{1}{\sqrt{\omega^2-u^2}}\ du \\
&= \frac{\pi R^2}{2}
\end{aligned}$$

$$\begin{aligned}
\left|\int_0^\omega I_2(u)\ du\right| &= \left|\int_0^\omega \frac{(3u^2+1)(\cos(Ru)-1)}{u^2(1+u^2)^2\sqrt{\omega^2-u^2}}\ du\right| \\
&\leq \int_0^\omega \frac{R^2(3u^2+1)}{(1+u^2)^2\sqrt{\omega^2-u^2}}\ du \\
&\leq 13R^2 \int_0^\omega \frac{1}{\sqrt{\omega^2-u^2}}\ du \\
&= \frac{13\pi R^2}{2}
\end{aligned}$$

From (2.23) we have $f(\omega) \geq \frac{1}{10\pi\sqrt{5}}$, and by (2.25) we conclude that for $\omega \leq 2$,

$$\begin{aligned}
\left|1-\frac{\hat{C}_1(\omega)}{f(\omega)}\right| &\leq 1-(1-e^{-R})^2 + 2e^{-R}\left(\frac{1}{f(\omega)}\left|\int_0^\omega I_1(u)\ du\right| + \frac{1}{f(\omega)}\left|\int_0^\omega I_2(u)\ du\right|\right) \\
&\leq 2e^{-R}\left(1+10\pi\sqrt{5}\left(\left|\int_0^\omega I_1(u)\ du\right| + \left|\int_0^\omega I_2(u)\ du\right|\right)\right) \\
&\leq e^{-R}(2+140\pi^2\sqrt{5}R^2) \qquad (2.37)
\end{aligned}$$

Combining (2.36) and (2.37), we obtain the desired bound for all ω:

$$\left|1-\frac{\hat{C}_1(\omega)}{f(\omega)}\right| \leq e^{-R}\left(2+140\pi^2\sqrt{5}R^2 + \frac{\pi^2}{2} + \frac{\pi^2\sqrt{10}}{3}(5R+20)\right)$$

□

Theorem 2.5 follows as a special case of Theorem 2.9, which we now prove:

Theorem 2.9. *Let V be a topological vector space, and let $Z(g)$ be a second-order stationary V-valued random field over an LCA group G, with covariance function $C(g)$ and spectral density $f(\chi)$, and let a compact subset $K \subseteq G$ be given. If there exists a bounded function $C_1(h)$ with support contained in K such that the Fourier transform $\hat{C}_1$ satisfies*

$$|\hat{C}_1(\chi)(\lambda_1,\lambda_2) - f(\chi)(\lambda_1,\lambda_2)| \leq \epsilon\sqrt{f(\chi)(\lambda_1,\lambda_1)f(\chi)(\lambda_2,\lambda_2)} \qquad (2.38)$$

for all $\lambda_1, \lambda_2 \in V'$ and all $\chi \in \hat{G}$, then $r^*(K) \leq \epsilon$.

Proof. Let A, B be finite subsets of $\mathbb{R}^n$ with $(A+K)\cap B = \emptyset$, and let λ_x for $x \in A$ and λ_y for $y \in B$ be given elements of V'. Define random variables

$$X = \sum_{x\in A} \lambda_x(Z(x)), \quad Y = \sum_{y\in B} \lambda_y(Z(y)).$$

Assuming without loss of generality that $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, we must show $|\operatorname{Cov}(X,Y)| \leq \epsilon$. In fact we have

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \operatorname{Cov}\left(\sum_{x\in A} \lambda_x(Z(x)), \sum_{y\in B} \lambda_y(Z(x))\right) \\
&= \sum_{x\in A, y\in B} \operatorname{Cov}(\lambda_x(Z(x)), \lambda_y(Z(y))) \\
&= \sum_{x\in A, y\in B} C(y-x)(\lambda_x, \lambda_y) \\
&= \sum_{x\in A, y\in B} \int_{\hat{G}} f(\chi)(\lambda_x, \lambda_y)\chi(y-x)\ d\chi \qquad (2.39)
\end{aligned}$$

Now the assumptions imply that C_1 and $\hat{C_1}$ are integrable, so we may apply Fourier inversion to obtain that for all $x \in A$ and $y \in B$,

$$\int_{\hat{G}} \hat{C_1}(\chi)(\lambda_x, \lambda_y)\chi(y-x)\ d\chi = C_1(y-x) = 0$$

since the support of C_1 is contained in K but the assumption $(A+K)\cap B = \emptyset$ implies that $y - x \notin K$. Applying this to (2.39), we have

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \sum_{x\in A, y\in B} \int_{\hat{G}} (f - \hat{C_1})(\chi)(\lambda_x, \lambda_y)\chi(y-x)\ d\chi \\
&= \sum_{x\in A, y\in B} \int_{\hat{G}} (f - \hat{C_1})(\chi)(\overline{\chi(x)}\lambda_x, \overline{\chi(y)}\lambda_y)\ d\chi \\
&= \int_{\hat{G}} (f - \hat{C_1})(\chi) \left(\sum_{x\in A} \overline{\chi(x)}\lambda_x, \sum_{y\in B} \overline{\chi(y)}\lambda_y\right)\ d\chi
\end{aligned}$$

By (2.6), this implies

$$|\operatorname{Cov}(X,Y)| \leq \int_{\hat{G}} \epsilon \sqrt{f(\chi)\left(\sum_{x\in A} \overline{\chi(x)}\lambda_x, \sum_{x\in A} \overline{\chi(x)}\lambda_x\right) f(\chi)\left(\sum_{y\in B} \overline{\chi(y)}\lambda_y, \sum_{y\in B} \overline{\chi(y)}\lambda_y\right)}\ d\chi \qquad (2.40)$$

On the other hand, we have

$$1 = \mathrm{Var}(X) = \operatorname{Cov}\left(\sum_{x\in A} \lambda_x(Z(x)), \sum_{y\in A} \lambda_y(Z(y))\right)$$

$$
\begin{aligned}
&= \sum_{x\in A}\sum_{y\in A} \operatorname{Cov}\left(\lambda_x(Z(x)), \lambda_y(Z(x))\right) \\
&= \sum_{x\in A}\sum_{y\in A} C(y-x)(\lambda_x, \lambda_y) \\
&= \sum_{x\in A}\sum_{y\in A} \int_{\hat{G}} f(\chi)(\lambda_x, \lambda_y)\chi(y-x)\ d\chi \\
&= \sum_{x\in A}\sum_{y\in A} \int_{\hat{G}} f(\chi)(\overline{\chi(x)}\lambda_x, \overline{\chi(y)}\lambda_y)\ d\chi \\
&= \int_{\hat{G}} f(\chi) \left(\sum_{x\in A} \overline{\chi(x)}\lambda_x, \sum_{x\in A} \overline{\chi(x)}\lambda_x \right)\ d\chi
\end{aligned}
$$

and similarly $\int_{\hat{G}} f(\chi) \left(\sum_{y\in B} \overline{\chi(y)}\lambda_y, \sum_{y\in B} \overline{\chi(y)}\lambda_y \right)\ d\chi = 1$, so applying the Cauchy-Schwarz inequality to (2.40), we conclude that $|\operatorname{Cov}(X,Y)| \leq \epsilon$, as desired. □

Theorem 2.10. *Let V be a Hilbert space, and let $Z(g)$ be a second-order stationary V-valued random field over an LCA group G, with symmetric covariance function $C(g)$ and positive definite spectral density $f(\chi)$, and let a compact subset $K \subseteq G$ be given. Given a bounded function $C_1(g)$ with support contained in K, we have*

$$
r^*(K) \leq \sup_{\chi\in\hat{G}} \operatorname{rad}\left(T_{f(\chi)}^{-1}(T_{\hat{C}_1(\chi)} - T_{f(\chi)}) \right), \tag{2.41}
$$

Proof. We recall that every positive operator T on a Hilbert space has a positive square root $T^{1/2}$, that the operator norm of a Hermitian operator equals its spectral radius, and that the spectral radius is invariant under conjugation (i.e., if T is an operator and S is an invertible operator, then $\operatorname{rad}(T) = \operatorname{rad}(STS^{-1})$). Using these facts, we have

$$
\begin{aligned}
&\sup\{|\hat{C}_1(\chi)(\lambda_1,\lambda_2) - f(\chi)(\lambda_1,\lambda_2)| : f(\chi)(\lambda_1,\lambda_1) \leq 1, f(\chi)(\lambda_2,\lambda_2) \leq 1\} \\
&= \sup\{|\langle (T_{\hat{C}_1} - T_{f(\chi)})\lambda_1, \lambda_2\rangle : \langle T_{f(\chi)}\lambda_1, \lambda_1\rangle \leq 1, \langle T_{f(\chi)}\lambda_2, \lambda_2\rangle \leq 1\} \\
&= \sup\{|\langle (T_{\hat{C}_1} - T_{f(\chi)})\lambda_1, \lambda_2\rangle : \langle T_{f(\chi)}^{1/2}\lambda_1, T_{f(\chi)}^{1/2}\lambda_1\rangle \leq 1, \langle T_{f(\chi)}^{1/2}\lambda_2, T_{f(\chi)}^{1/2}\lambda_2\rangle \leq 1\} \\
&= \sup\{|\langle (T_{\hat{C}_1} - T_{f(\chi)})T_{f(\chi)}^{-1/2}\lambda_1, T_{f(\chi)}^{-1/2}\lambda_2\rangle : \langle \lambda_1, \lambda_1\rangle \leq 1, \langle \lambda_2, \lambda_2\rangle \leq 1\} \\
&= \sup\{|\langle T_{f(\chi)}^{-1/2}(T_{\hat{C}_1} - T_{f(\chi)})T_{f(\chi)}^{-1/2}\lambda_1, \lambda_2\rangle : \langle \lambda_1, \lambda_1\rangle \leq 1, \langle \lambda_2, \lambda_2\rangle \leq 1\} \\
&= \|T_{f(\chi)}^{-1/2}(T_{\hat{C}_1} - T_{f(\chi)})T_{f(\chi)}^{-1/2}\| \\
&= \operatorname{rad}(T_{f(\chi)}^{-1/2}(T_{\hat{C}_1} - T_{f(\chi)})T_{f(\chi)}^{-1/2}) \\
&= \operatorname{rad}\left(T_{f(\chi)}^{-1}(T_{\hat{C}_1} - T_{f(\chi)}) \right)
\end{aligned}
$$

Given arbitrary nonzero $\lambda_1, \lambda_2 \in V'$, by replacing λ_1 with $\lambda_1/\sqrt{f(\chi)(\lambda_1,\lambda_1)}$ and λ_2 with $\lambda_2/\sqrt{f(\chi)(\lambda_2,\lambda_2)}$, we conclude that

$$
|\hat{C}_1(\chi)(\lambda_1,\lambda_2) - f(\chi)(\lambda_1,\lambda_2)| \leq \operatorname{rad}((T_{\hat{C}_1} - T_{f(\chi)})T_{f(\chi)}^{-1})\sqrt{f(\chi)(\lambda_1,\lambda_1)f(\chi)(\lambda_2,\lambda_2)}
$$

On the other hand, if either λ_1 or λ_2 are zero, then this inequality holds trivially, so in fact it holds for all $\lambda_1, \lambda_2 \in V'$. Applying Theorem 2.9, the result follows. □

Theorem 2.11. *Let $Z(s)$ be a second-order stationary $\mathbb{R}^2$-valued random field on $\mathbb{R}$ with covariance function given by*

$$M_{C(h)} = \begin{pmatrix} e^{-x} & ce^{-bx} \\ ce^{-bx} & e^{-ax} \end{pmatrix}$$

where $a > 0, b > 0$, and

$$c^2 < \min\left\{\frac{a}{b^2}, \frac{b^2}{a}, \frac{2a}{1+a^2}\right\} \tag{2.42}$$

Then $Z(s)$ is r^*-mixing. More precisely, if we set $K_1 = \min\{1, a, b\}$, then there is some constant L, depending only on a, b, and c, such that

$$r^*(R) \leq L(R+1)e^{-KR}$$

Proof. First we show that $M_{f(\omega)} = M_{\hat{C}(\omega)}$ is positive definite. We calculate

$$M_{f(\omega)} = \begin{pmatrix} \dfrac{1}{2\pi(1+\omega^2)} & \dfrac{bc}{2\pi(b^2+\omega^2)} \\ \dfrac{bc}{2\pi(b^2+\omega^2)} & \dfrac{a}{2\pi(a^2+\omega^2)} \end{pmatrix}$$

To show that $f(\omega)$ is positive definite, it suffices to show that its trace and determinant are positive. For the trace, this is clear, while for the determinant we calculate

$$\begin{aligned} \det M_{f(\omega)} &= \frac{a}{4\pi^2(1+\omega^2)(a^2+\omega^2)} - \frac{b^2c^2}{4\pi^2(b^2+\omega^2)^2} \\ &= \frac{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}{4\pi^2(1+\omega^2)(a^2+\omega^2)(b^2+\omega^2)^2} \end{aligned}$$

Here the denominator is positive, while if we let $v(\omega^2)$ denote the numerator, we apply (2.42) to obtain

$$\begin{aligned} v(0) &= ab^4 - b^2c^2a^2 = ab^2(b^2 - ac^2) > 0 \\ v'(0) &= 2ab^2 - b^2c^2(1+a^2) = b^2(2a - c^2(1+a^2)) > 0 \\ v''(\omega^2) &= 2(a - b^2c^2) > 0 \end{aligned}$$

so that $v(\omega^2) > 0$ for all ω, as required.

Now define a form $C_1(h)$ by

$$M_{C_1(h)} = \begin{pmatrix} e^{-x}(1-e^{x-R}) & ce^{-bx}(1-e^{b(x-R)}) \\ ce^{-bx}(1-e^{b(x-R)}) & e^{-ax}(1-e^{a(x-R)}) \end{pmatrix}$$

Calculating as in Theorem 2.6, we have

$$M_{\hat{C}_1(\omega)} = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$$

where

$$\begin{aligned} c_{11} &= \frac{1}{2\pi(1+\omega^2)}\left(1 - e^{-R}\left(\frac{\sin(\omega R)}{\omega} + \cos(\omega R)\right)\right) \\ c_{12} &= \frac{bc}{2\pi(b^2+\omega^2)}\left(1 - e^{-bR}\left(\frac{b\sin(\omega R)}{\omega} + \cos(\omega R)\right)\right) \\ c_{22} &= \frac{a}{2\pi(a^2+\omega^2)}\left(1 - e^{-aR}\left(\frac{a\sin(\omega R)}{\omega} + \cos(\omega R)\right)\right) \end{aligned}$$

We then have

$$M_{\hat{C}_1(\omega)} = M_{f(\omega)} + \begin{pmatrix} -\frac{1}{2\pi(1+\omega^2)}\epsilon_1 & -\frac{bc}{2\pi(b^2+\omega^2)}\epsilon_2 \\ -\frac{bc}{2\pi(b^2+\omega^2)}\epsilon_2 & -\frac{a}{2\pi(a^2+\omega^2)}\epsilon_3 \end{pmatrix}$$

where

$$\begin{aligned} \epsilon_1 &= e^{-R}\left(\frac{\sin(\omega R)}{\omega} + \cos(\omega R)\right) \\ \epsilon_2 &= e^{-bR}\left(\frac{b\sin(\omega R)}{\omega} + \cos(\omega R)\right) \\ \epsilon_3 &= e^{-aR}\left(\frac{a\sin(\omega R)}{\omega} + \cos(\omega R)\right) \end{aligned}$$

Here we have bounds

$$|\epsilon_1| \le (R+1)e^{-R}, \quad |\epsilon_2| \le (bR+1)e^{-bR}, \quad |\epsilon_3| \le (aR+1)e^{-aR} \tag{2.43}$$

Setting $\Delta = \det M_{f(\omega)}$, we have we calculate

$$\begin{aligned} M_{f(\omega)}^{-1}(M_{\hat{C}_1} - M_{f(\omega)}) &= \frac{1}{\Delta}\begin{pmatrix} \frac{a}{2\pi(a^2+\omega^2)} & -\frac{bc}{2\pi(b^2+\omega^2)} \\ -\frac{bc}{2\pi(b^2+\omega^2)} & \frac{1}{2\pi(1+\omega^2)} \end{pmatrix}\begin{pmatrix} -\frac{1}{2\pi(1+\omega^2)}\epsilon_1 & -\frac{bc}{2\pi(b^2+\omega^2)}\epsilon_2 \\ -\frac{bc}{2\pi(b^2+\omega^2)}\epsilon_2 & -\frac{a}{2\pi(a^2+\omega^2)}\epsilon_3 \end{pmatrix} \\ &= \begin{pmatrix} c'_{11} & c'_{12} \\ c'_{21} & c'_{22} \end{pmatrix} \end{aligned}$$

where

$$c'_{11} = \frac{1}{\Delta}\left(\frac{-a}{4\pi^2(a^2+\omega^2)(1+\omega^2)}\epsilon_1 + \frac{b^2c^2}{4\pi^2(b^2+\omega^2)^2}\epsilon_2\right)$$

$$= -\frac{a(b^2+\omega^2)^2\epsilon_1 - b^2c^2(1+\omega^2)(a^2+\omega^2)\epsilon_2}{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}$$

$$c'_{12} = \frac{1}{\Delta}\left(\frac{-abc}{4\pi^2(a^2+\omega^2)(b^2+\omega^2)}\epsilon_2 + \frac{abc}{4\pi^2(b^2+\omega^2)(a^2+\omega^2)}\epsilon_3\right)$$
$$= \frac{abc(1+\omega^2)(b^2+\omega^2)}{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}(\epsilon_3 - \epsilon_2)$$

$$c'_{21} = \frac{1}{\Delta}\left(\frac{bc}{4\pi^2(1+\omega^2)(b^2+\omega^2)}\epsilon_1 + \frac{-bc}{4\pi^2(1+\omega^2)(b^2+\omega^2)}\epsilon_2\right)$$
$$= \frac{bc(a^2+\omega^2)(b^2+\omega^2)}{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}(\epsilon_1 - \epsilon_2)$$

$$c'_{22} = \frac{1}{\Delta}\left(\frac{b^2c^2}{4\pi^2(b^2+\omega^2)^2}\epsilon_2 + \frac{-a}{4\pi^2(1+\omega^2)(a^2+\omega^2)}\epsilon_3\right)$$
$$= \frac{b^2c^2(1+\omega^2)(a^2+\omega^2)\epsilon_2 - a(b^2+\omega^2)^2\epsilon_3}{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}$$

Let $K_1 = \min\{1, a, b\}$ and $K_2 = \max\{1, a, b\}$. Then, by (2.43) we have a bound

$$|c'_{11}| \leq \frac{a(b^2+\omega^2)^2 + b^2c^2(1+\omega^2)(a^2+\omega^2)}{a(b^2+\omega^2)^2 - b^2c^2(1+\omega^2)(a^2+\omega^2)}(K_2R+1)e^{-K_1R}$$

Here the denominator here never vanishes (since $\Delta \neq 0$), and the numerator and denominator are both polynomials of degree two in ω^2, giving a bound of the form

$$|c'_{11}| \leq K_3(K_2R+1)e^{-K_1R}$$

for a suitable constant K_3 (depending only on a, b, and c). By the same argument, we obtain bounds of the same form for c'_{12}, c'_{21}, and c'_{22}, so by redefining K_3 if necessary, we obtain bounds

$$|c'_{ij}| \leq K_3(K_2R+1)e^{-K_1R}$$

for $i, j = 1, 2$. Since the spectral radius is bounded by the maximum row sum, we deduce that rad $\left(M^{-1}_{f(\chi)}(M_{\hat{C}_1(\chi)} - M_{f(\chi)})\right) \leq 2K_3(K_2R+1)e^{-K_1R}$, and the conclusion follows from Theorem 2.10. □

CHAPTER 3

EMPIRICAL SEMIVARIOGRAM

As discussed in Chapter 1, semivariogram estimation plays an essential role in spatial statistics. Given an intrinsically stationary random field, consistent estimation of the semivariogram enables asymptotically optimal prediction of the random field at unobserved locations, in the sense of minimizing the mean-squared error. We have already reviewed the published results on the consistency of the empirical semivariogram and related estimators. In this chapter, we present a new asymptotic framework, applicable to both systematic and stochastic sampling designs, under which the semivariogram may be consistently estimated.

In §3.1, we give an example demonstrating that excessive clustering of the observation locations may lead to inconsistency of the empirical semivariogram, even if the observation locations approach a dense subset. This motivates the assumptions which we introduce in §3.2 and which we show are sufficient to enable consistent estimation of the semivariogram pointwise. In §3.3, we introduce stronger assumptions which enable consistent estimation of the semivariogram $\gamma(h)$ uniformly for all h in any given compact set. In §3.4, we extend these results to multivariate random fields.

3.1 Failure of consistency: clustering

Let $Z(s)$ be the stationary Gaussian process in $\mathbb{R}$ with exponential covariance function $C(h) = e^{-h/\theta}$. The semivariogram is thus $\gamma(h) = 1 - e^{-h/\theta}$. For each $k \geq 0$, define a set of locations $S_k \subseteq \mathbb{R}$:

$$S_k = P_k \cup Q_k$$

where

$$P_k = \left\{ \frac{i}{2^k} : i \in \mathbb{Z}, -4^k \leq i \leq 4^k \right\}$$
$$Q_k = \left\{ \frac{i}{4^k} : i \in \mathbb{Z}, 0 \leq i \leq 4^k \right\}$$

Figure 3.1 illustrates the set S_2. Note that the set S_k increases with k, $\bigcup_{k=0}^{\infty} S_k$ is a dense subset of $\mathbb{R}$, and for each k, roughly a third of the points lie in the cluster $Q_k \subseteq [0,1]$. Given any $h = \frac{i}{2^j}$, for $j \geq 0$ and $0 < i < 2^j$, it can be shown that $\hat{\gamma}(h)$ is not a consistent

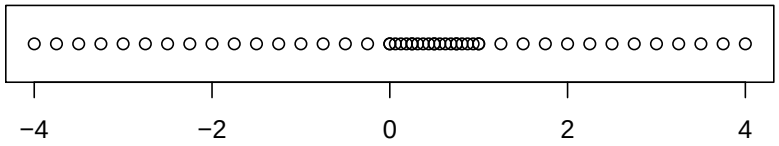

Figure 3.1: The set of observation locations S_k in §3.1, for $k = 2$.

estimator of $\gamma(h)$; for a proof, see the end of the chapter. Figure 3.2 summarizes simulations illustrating the failure of convergence for $h = 0.5$.

In this example, the consistency of the estimator may be recovered by discarding observations from the heavily clustered region. This property, that including additional observations may worsen the estimator, is an undesirable feature. It underscores an inefficiency of the empirical semivariogram and suggests that alternative forms of semivariogram estimation should possibly be considered if the observation locations are heavily clustered.

3.2 Consistency of empirical semivariogram

To establish the consistency of $\hat{\gamma}_\delta(h)$, we need to assume a lack of excessive clustering in the pairs of observation locations with distance approximately h. For our asymptotic analysis, we assume a sequence of finite subsets $S_1, S_2, S_3, \ldots$ of $\mathbb{R}^n$, where S_k represents the kth set of observation locations. In what follows, we will use the notation $d(A, B) = \sup_{a \in A, b \in B} |a - b|$ for the distance between two sets A and B. To make precise the required "lack of excessive clustering", we make the following definition: given $h \geq 0$ and $\delta > 0$, and given location sets S_k, we define $P_k = P_k(h, \delta)$ to be the set of pairs of points separated by a distance between $h - \delta$ and $h + \delta$:

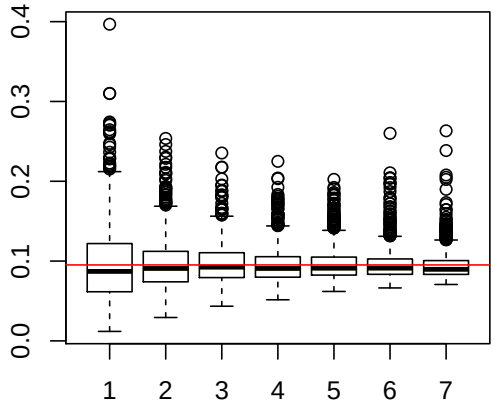

Figure 3.2: Boxplots showing the distribution of $\hat{\gamma}(0.5)$, for $k = 2, 3, 4, 5, 6, 7$, based on 1000 simulations each.

$$P_k = P_k(h, \delta) = \{\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq S_k : h - \delta < \|\boldsymbol{s} - \boldsymbol{t}\| < h + \delta\}.$$

We want to ensure that, in some sense, points in P_k are not clustering too much in bounded regions. Given a Euclidean ball $U \subset \mathbb{R}^n$, define

$$p_k(U) = \text{the proportion of pairs } \{\boldsymbol{s}, \boldsymbol{t}\} \in P_k \text{ with at least one of } \boldsymbol{s} \text{ or } \boldsymbol{t} \text{ in } U$$

For a given radius r, define

$$p_k(r) = p_k(r; h, \delta) = \sup\{p_k(U) \mid U \text{ is ball of radius no more than } r\}$$

We may now formulate our assumption of non-clustering:

Assumption 3.1. For each $h \geq 0$, $\delta > 0$, $r > 0$, we have $\lim_{k\to\infty} p_k(r; h, \delta) = 0$.

Note that the definition implicitly requires p_k to be well-defined for all sufficiently large k, meaning that P_k is nonempty, i.e., there is at least one pair of points in $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k$ with $h - \delta < \|\boldsymbol{s}_i - \boldsymbol{s}_j\| < h + \delta$. We will also need the following assumptions about the random field $Z(\boldsymbol{s})$:

Assumption 3.2. $Z(\boldsymbol{s})$ is a second-order stationary, isotropic random field.

Assumption 3.3. $Z(\boldsymbol{s})$ has a continuous semivariogram $\gamma(h)$.

Assumption 3.4. $Z(\boldsymbol{s})$ has finite fourth-order moments.

Assumption 3.5. $Z(\boldsymbol{s})$ is ρ^*-mixing.

With these assumptions in place, we can establish the following theorem:

Theorem 3.1. *Suppose Assumptions 3.1–3.5 are satisfied. If C_k and δ_k are sequences of positive numbers such that $\rho^*(C_k) \to 0$, $\delta_k \to 0$, and $p_k(h + C_k + \delta_k; h, \delta_k) \to 0$, then $\hat{\gamma}_{\delta_k}(h)$ is consistent; i.e., $\hat{\gamma}_{\delta_k}(h)$ converges to $\gamma(h)$ in probability.*

If Assumptions 3.1–3.5 are satisfied, then suitable sequences C_k and δ_k for Theorem 3.1 always exist (see Theorem 3.8). However, to obtain δ_k constructively, we will need to strengthen our assumption on the observation locations. For this we will use partitions of the space $\mathbb{R}^n$ into grids of hypercubes, making the following definition: given a real number $\epsilon > 0$, an *ϵ-cube* is a set of the form

$$\prod_{i=1}^{n} [\epsilon a_i, \epsilon(a_i + 1)]$$

where $a_1, \ldots, a_n$ are integers and where $\prod$ denotes the Cartesian product. Thus the ϵ-cubes essentially provide a partition of the space $\mathbb{R}^n$ into cubes of volume ϵ^n, overlapping only on their boundaries. We also introduce the following notation to denote open and closed neighborhoods of a set A:

$$B_A(r) = \{\boldsymbol{x} \in \mathbb{R}^n : d(\boldsymbol{x}, A) < r\}$$
$$\overline{B}_A(r) = \{\boldsymbol{x} \in \mathbb{R}^n : d(\boldsymbol{x}, A) \leq r\}$$

In the case where A is a single point $A = \{\boldsymbol{a}\}$, we will write $B_{\boldsymbol{a}}(r)$ in place of $B_{\{\boldsymbol{a}\}}(r)$, denoting the open Euclidean ball of radius r centered at $\boldsymbol{a}$.

Assumption 3.6.

(a) There is an increasing sequence of bounded open sets $V_1, V_2, \ldots$ with $S_k \subseteq V_k$ and $\bigcup_{k=1}^{\infty} V_k = \mathbb{R}^n$.

(b) There is a constant $K > 0$ and a sequence of numbers $r_k > 0$ and points $\boldsymbol{x}_k \in \mathbb{R}^n$, such that $B_{\boldsymbol{x}_k}(r_k/K) \subseteq V_k \subseteq B_{\boldsymbol{x}_k}(r_k K)$.

(c) There are constants A_1, A_2, M and a sequence ϵ_k with $\epsilon_k \leq \frac{\delta_k}{3\sqrt{n}}$, such that for all $k \geq M$, every ϵ_k-cube contains at most $A_2|S_k|\epsilon_k^n/\lambda(V_k)$ points of S_k, and every ϵ_k-cube contained in V_k contains at least $A_1|S_k|\epsilon_k^n/\lambda(V_k)$ points of S_k, where $\lambda(V_k)$ is the Lebesgue measure of V_k, and $|S_k|$ is the number of points in the set S_k.

The main content of Assumption 3.6 is in part (c), which states, roughly speaking, that within the domain V_k the number of observation locations within each ϵ_k-cube differs by at most a constant factor from the number which would be expected if the points were uniformly distributed. Under this assumption, we obtain a constructive version of Theorem 3.1:

Theorem 3.2. *If Assumption 3.6 is satisfied and $\delta_k \to 0$, then any sequence $C_k \to \infty$ with $\frac{C_k}{r_k} \to 0$ satisfies the criteria of Theorem 3.1. In particular, if $\delta_k \to 0$, and Assumptions 3.2–3.6 hold, then $\hat{\gamma}_{\delta_k}(h)$ is consistent.*

Finally, we consider the case where observation locations S_k are chosen randomly on bounded domains V_k; in this case, we will show that under fairly general conditions the locations will almost surely satisfy Assumption 3.6(c), provided parts (a) and (b) are satisfied. In the literature, it is common to assume that the domains V_k are scaled copies of a fixed domain V, i.e., $V_k = c_k V$ with $c_k \to \infty$ (e.g., see [17, 27, 14]). However, this is not

essential, and we allow somewhat greater flexibility in our assumptions about the domains V_k.

In the following theorem, the location sets S_k will be chosen to contain a certain number of points m_k, and we will need to assume that m_k grows sufficiently fast to ensure that the expected number of points of S_k in an ϵ_k-cube approaches infinity at a suitable rate:

Assumption 3.7. $\lim_{k\to\infty} \dfrac{m_k \epsilon_k^n / \lambda(V_k)}{\log m_k} = \infty.$

Theorem 3.3. *Suppose there are constants $0 < A_1' \leq A_2' < \infty$ and measurable functions $f_k : V_k \to \mathbb{R}$, with $A_1' \leq f_k \leq A_2'$, and suppose Assumptions 3.6(a,b) and 3.7 are satisfied. Suppose $S_1, S_2, \ldots,$ are random sets where S_k is distributed according to an inhomogeneous binomial point process with intensity $f_k(x)/\int_{V_k} f_k \, d\lambda$ and $|S_k| = m_k$. Then for any choice of sequence δ_k satisfying $\epsilon_k \leq \frac{\delta_k}{3\sqrt{n}}$, with probability 1 Assumption 3.6(c) is also satisfied.*

Here the meaning of the term "inhomogeneous binomial point process" is that S_k may be written as

$$S_k = \{X_{k1}, X_{k2}, \ldots, X_{km_k}\},$$

for some random variables $X_{k1}, \ldots, X_{km_k}$ each of which have probability density function $f_k(x)/\int_{V_k} f_k \, d\lambda$, such that for each fixed k, $X_{k1}, \ldots, X_{km_k}$ are independent.

3.3 Uniform consistency

In this section, we will give conditions under which $\hat{\gamma}_\delta(h)$ converges uniformly on bounded intervals $[0, h_0]$. Uniform convergence on the whole domain $[0, \infty)$ is of course impossible since $\hat{\gamma}_\delta(h)$ is undefined except for h in a bounded subset: if h_1 is the maximum distance between any pair of points in S_k, then $\hat{\gamma}_\delta(h)$ is undefined for all h in $(h_1 + \delta, \infty)$. We begin with the following set of technical conditions, which ensure that the bandwidth δ_k does not decay too quickly, relative to the rate of decay of the ρ^*-mixing coefficients and the grid widths ϵ_k:

Assumption 3.8. There are increasing sequences C_k and β_k and a decreasing sequence $\tau_k \to 0$ such that

(a) $\lim_{k\to\infty} \dfrac{\rho^*(C_k) + \left(\frac{C_k}{r_k}\right)^n}{\tau_k} = 0.$

(b) $\lim_{k\to\infty} |S_k| P(|Z(\boldsymbol{s})| > \beta_k) = 0.$

(c) $\lim_{k\to\infty} \beta_k^2 \epsilon_k / \delta_k = 0.$

(d) $\lim_{k\to\infty} \beta_k^2 \tau_k/\delta_k = 0$.

Under these conditions, we obtain the following result of uniform convergence:

Theorem 3.4. *If Assumptions 3.2–3.6 and 3.8 are satisfied, then for all $h_0 > 0$, $\hat{\gamma}_{\delta_k}(h)$ converges uniformly to $\gamma(h)$ in probability on $[0, h_0]$, i.e., for any $\epsilon > 0$,*

$$\lim_{k\to\infty} P\left(\sup_{h\in[0,h_0]} |\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon\right) = 0$$

If we assume that the relevant sequences grow at a polynomial rate, then Assumption 3.8 may be expressed somewhat more explicitly. For this we will use the following asymptotic notation: given sequences $a_k \geq 0$ and $b_k > 0$, we write $a_k \ll b_k$ if $\limsup_{k\to\infty} \frac{a_k}{b_k} < \infty$; we write $a_k \asymp b_k$ if both $a_k \ll b_k$ and $b_k \ll a_k$.

Theorem 3.5. *Suppose that $E|Z(\boldsymbol{s}) - \mu|^p < \infty$, and that $\rho^*(h) \ll h^{-\alpha_\rho}$, $r_k \asymp k^{\alpha_r}$, $|S_k| \asymp k^{\alpha_m}$, $\delta_k \asymp k^{-\alpha_\delta}$, $\epsilon_k \asymp k^{-\alpha_\epsilon}$, $C_k \asymp k^{\alpha_C}$, $\tau_k \asymp k^{-\alpha_\tau}$, and $\beta_k \asymp k^{\alpha_\beta}$ for some constants $p, \alpha_\rho, \alpha_r, \alpha_m, \alpha_\delta, \alpha_\epsilon, \alpha_C, \alpha_\tau, \alpha_\beta > 0$. Then Assumption 3.8 holds provided that*

(a) $\alpha_\tau < \min\{\alpha_C \alpha_\rho, n(\alpha_r - \alpha_C)\}$,

(b) $\alpha_m < \alpha_\beta p$,

(c) $2\alpha_\beta < \alpha_\epsilon - \alpha_\delta$,

(d) $2\alpha_\beta < \alpha_\tau - \alpha_\delta$.

As the sequences ϵ_k, C_k, τ_k, β_k serve only an auxiliary purpose in Theorem 3.8, having no role in the construction of the estimator $\hat{\gamma}_\delta$, it is desirable to obtain sufficient conditions not involving them. This may be done as follows:

Theorem 3.6. *Suppose that $E|Z(\boldsymbol{s}) - \mu|^p < \infty$ for some $p > 2n$, and that $\rho^*(h) \ll h^{-\alpha_\rho}$, $r_k \asymp k^{\alpha_r}$, $m_k \asymp k^{\alpha_m}$, $\delta_k \asymp k^{-\alpha_\delta}$, for some constants $\alpha_\rho, \alpha_r, \alpha_m, \alpha_\delta > 0$, and suppose further that*

(a) $\dfrac{\alpha_m}{n\alpha_r} < \dfrac{p}{2}\left(\dfrac{\alpha_\rho}{\alpha_\rho + n} - \dfrac{\alpha_\delta}{n\alpha_r}\right)$

(b) $\alpha_\delta < \alpha_m\left(\dfrac{1}{n} - \dfrac{2}{p}\right) - \alpha_r$

Then constants $\alpha_\epsilon, \alpha_C, \alpha_\tau, \alpha_\beta$ exist which satisfy Assumption 3.7 as well as the assumptions of Theorem 3.5, if locations sets S_k are chosen with $|S_k| = m_k$.

Remark: If $Z(\boldsymbol{s})$ has moments of all orders, then p may be taken arbitrarily large so that (a) is equivalent to $\alpha_\delta < n\alpha_r\alpha_\rho/(\alpha_\rho + n)$. If in addition $\rho^*(h)$ decays faster than any monomial (e.g., exponentially fast), then α_ρ may be taken arbitrarily large, in which case (a) is equivalent to $\alpha_\delta < n\alpha_r$.

Example 3.1. Let $Z(\boldsymbol{s})$ be a Gaussian process on $\mathbb{R}^2$ with an isotropic semivariogram $\gamma(h) = 1 - e^{-h/10}$. Let the observation set S_k be a square $k^2 \times k^2$ grid,

$$S_k = \{(i/k, j/k) \mid i = -k^2, \ldots, k^2, j = -k^2, \ldots, k^2\},$$

and let $\delta_k = 1/\sqrt{k}$. Then S_k is contained in the ball V_k centered at the origin with radius Dk, for an arbitrary constant $D > \sqrt{2}$. By Example 2.7, Z is ρ^*-mixing, and we can take α_ρ arbitrarily large. Since $Z(\boldsymbol{s})$ has a Gaussian distribution, which has moments of all orders, we may also take p arbitrarily large, and Theorem 3.6 applies with $n = 2, \alpha_m = 4, \alpha_r = 1, \alpha_\delta = 1/2$. The convergence of γ_{δ_k} in this case is illustrated in Figure 3.3.

3.4 Multivariate random fields

In the previous section, we considered only real-valued random fields $Z(\boldsymbol{s})$. However, many applications involve the observation of a multivariate random field, in which each location corresponds to a vector $\boldsymbol{Z}(\boldsymbol{s}) \in \mathbb{R}^p$. For example, a mining engineer may be interested in the concentration not just of gold but of p different types of ore. At any given location, these concentrations may be correlated with one another as well as with the concentrations at nearby locations.

We will let $Z_1(\boldsymbol{s}), \ldots, Z_p(\boldsymbol{s})$ denote the components of the multivariate random field $\boldsymbol{Z}(\boldsymbol{s})$. If $\boldsymbol{Z}(\boldsymbol{s})$ is isotropic, the dependence structure may be described by the *cross-semivariogram*:

$$\gamma_{ij}(\|\boldsymbol{h}\|) = \frac{1}{2}E(Z_i(\boldsymbol{s}+\boldsymbol{h}) - Z_i(\boldsymbol{s}))(Z_j(\boldsymbol{s}+\boldsymbol{h}) - Z_j(\boldsymbol{s}))$$

If $\boldsymbol{Z}(\boldsymbol{s})$ is observed at regularly-spaced locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k$, a natural estimator for $\gamma_{ij}(\boldsymbol{h})$ is the *empirical cross-semivariogram*:

$$\hat{\gamma}_{ij}(h) = \frac{1}{2N(h)}\sum(Z_i(\boldsymbol{s}_1) - Z_i(\boldsymbol{s}_2))(Z_j(\boldsymbol{s}_1) - Z_j(\boldsymbol{s}_2)) \tag{3.1}$$

where the sum ranges over pairs of observation locations $\boldsymbol{s}_1, \boldsymbol{s}_2$ with $\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| = h$. and $N(h)$ is the number of such pairs. If the observation locations are irregularly spaced, then a variant of the empirical cross-semivariogram $\hat{\gamma}_{ij,\delta}$ with a bandwidth $\delta > 0$ may be defined just as in the univariate case.

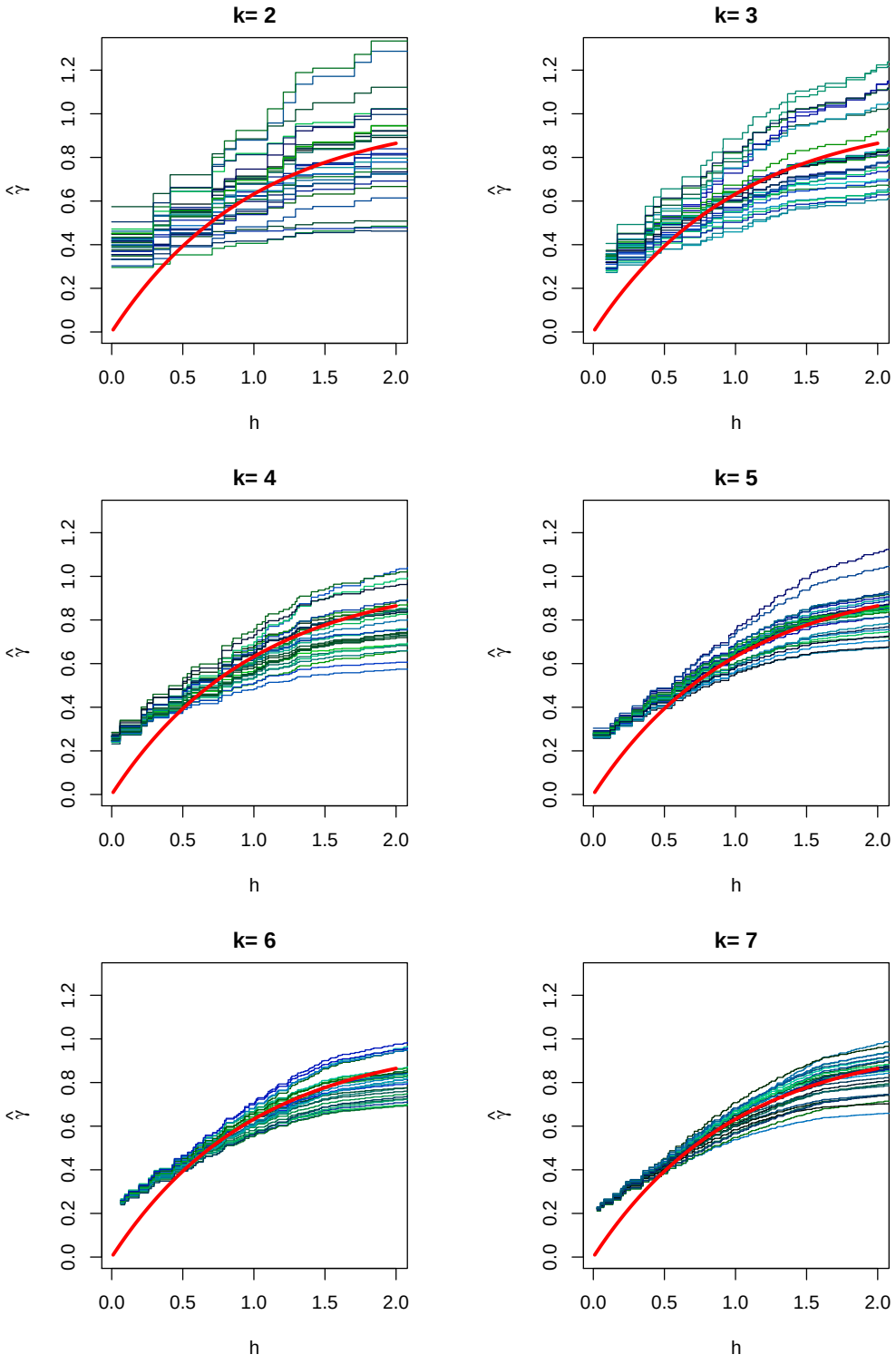

Figure 3.3: Simulations of $\hat{\gamma}_{\delta_k}(h)$, $0 \leq h \leq 2$, from Example 3.1, for $k = 2, 3, 4, 5, 6, 7$.

We will prove that $\hat{\gamma}_{ij,\delta}(h)$ is a consistent estimator for $\gamma_{ij}(h)$ under similar conditions to those assumed in the univariate case. The assumptions from §3.2 are adapted to the multivariate case as follows:

Assumption 3.9. $\boldsymbol{Z}(\boldsymbol{s})$ is a second-order stationary, isotropic random field.

Assumption 3.10. $\boldsymbol{Z}(\boldsymbol{s})$ has continuous cross-semivariograms $\gamma_{ij}(h)$.

Assumption 3.11. $Z_i(\boldsymbol{s})$ has finite fourth-order moments, for each $i = 1, \ldots, p$.

Assumption 3.12. $\boldsymbol{Z}(\boldsymbol{s})$ is ρ^*-mixing, in the sense of §2.5.

We then obtain the following generalization of Theorem 3.1:

Theorem 3.7. *Suppose that Assumptions (3.1) and (3.9)–(3.12) are satisfied. If C_k and δ_k are sequences of positive numbers such that $\rho^*(C_k) \to 0$, $\delta_k \to 0$, and $p_k(h+C_k+\delta_k; h, \delta_k) \to 0$, then $\hat{\gamma}_{ij,\delta_k}(h)$ is consistent.*

Once Theorem 3.7 is established, it is evident that Theorems 3.2 and 3.3 can be applied to the multivariate case as well, in order to extend the result to stochastic sampling designs.

3.5 Proofs

In the example of §3.1, given $h = \frac{i}{2^j}$ with $0 < i < 2^j$, we will prove by contradiction that $\hat{\gamma}(h)$ is an inconsistent estimator of $\gamma(h)$. We observe that for any $\epsilon < h$,

$$\begin{aligned}
&\operatorname{Cov}([Z(s+h) - Z(s)]^2, [Z(s+h+\epsilon) - Z(s+\epsilon)]^2) \\
&\qquad = 2\operatorname{Cov}(Z(s+h) - Z(s), Z(s+h+\epsilon) - Z(s+\epsilon))^2 \\
&\qquad = 2(2C(\epsilon) - C(h+\epsilon) - C(h-\epsilon))^2. \qquad (3.2)
\end{aligned}$$

The expression in (3.2) is a continuous function of ϵ which, as $\epsilon \to 0$, approaches the positive number $8(C(0)-C(h))^2$. Thus there are positive numbers ϵ_0 and K such that the expression in (3.2) is at least K for all $\epsilon \leq \epsilon_0$.

Now we have

$$\operatorname{Var}(\hat{\gamma}(h)) = \frac{1}{4N(h)^2} \sum \operatorname{Cov}([Z(s+h) - Z(s)]^2, [Z(t+h) - Z(t)]^2),$$

where the sum is taken over all $s, t \in S_k$ such that $s+h, t+h \in S_k$. Now, for large k, the proportion of such pairs (s,t) with $|t-s| < \epsilon_0$ approaches $(1-h)\epsilon_0/3$. Therefore,

$$\liminf_{k\to\infty} \operatorname{Var}(\hat{\gamma}(h)) \geq \frac{K}{4}(1-h)\epsilon_0/3 > 0. \qquad (3.3)$$

Choose some $p > 1$. Writing $\|\cdot\|_p$ to denote $[E|\cdot|^p]^{1/p}$, by Minkowski's inequality we have

$$\begin{aligned}\|\hat{\gamma}(h)\|_p &= \left\| \frac{1}{2N(h)} \sum (Z(s+h) - Z(s))^2 \right\|_p \\ &\leq \frac{1}{2N(h)} \sum \left\| (Z(s+h) - Z(s))^2 \right\|_p \\ &= \frac{1}{2N(h)} \sum [E(Z(s+h) - Z(s))^{2p}]^{1/p} \\ &= [E(Z(s+h) - Z(s))^{2p}]^{1/p} < \infty,\end{aligned}$$

where in the last step we use the fact that $Z(s) - Z(s+h)$ is Gaussian and hence has moments of all orders. This implies that $|\hat{\gamma}(h)|^2$ is L^p-bounded and hence uniformly integrable (as a sequence of random variables, for fixed h), so by the Vitali convergence theorem [22, Proposition 4.12] if $\hat{\gamma}(h)$ were consistent, then it would converge in L^2, and we would have $\mathrm{Var}(\hat{\gamma}(h)) \to 0$ as $k \to \infty$. But this would contradict (3.3), so $\hat{\gamma}(h)$ is an inconsistent estimator as claimed. This completes the proof.

Theorem 3.1. *Suppose Assumptions 3.1–3.5 are satisfied. If C_k and δ_k are sequences of positive numbers such that $\rho^*(C_k) \to 0$, $\delta_k \to 0$, and $p_k(h + C_k + \delta_k; h, \delta_k) \to 0$, then $\hat{\gamma}_{\delta_k}(h)$ is consistent; i.e., $\hat{\gamma}_{\delta_k}(h)$ converges to $\gamma(h)$ in probability.*

Proof. Let $L = E(Z(s)^4)$, so $L < \infty$ by Assumption 3.4. Fix $\epsilon > 0$ and $\rho > 0$, and assume that k is sufficiently large so that $p_k(h + C_k + \delta_k; h, \delta_k) < \epsilon$ and $\rho^*(C_k) < \rho$, by Assumptions 3.1 and 3.5, respectively. We note that for any $\boldsymbol{s}, \boldsymbol{t} \in S_k$, we have $E(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^4 \leq 32L$. Given a pair of points $\boldsymbol{s}, \boldsymbol{t} \in S_k$ with $h - \delta_k < \|\boldsymbol{s} - \boldsymbol{t}\| < h + \delta_k$, taking $U = B_{\boldsymbol{s}}(h + C_k + \delta_k)$, we have $\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq U$, and since $p_k(h + C_k + \delta_k; h, \delta_k) < \epsilon$ we know that at least proportion $1 - \epsilon$ of the pairs $\boldsymbol{s}', \boldsymbol{t}'$ with $h - \delta < |\boldsymbol{s}' - \boldsymbol{t}'| < h + \delta$ have $d(\{\boldsymbol{s}, \boldsymbol{t}\}, \{\boldsymbol{s}', \boldsymbol{t}'\}) \geq C_k$, and for these, we may bound the covariance

$$\begin{aligned}&\mathrm{Cov}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) \\ &\qquad = \mathrm{Corr}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) \\ &\qquad\quad \cdot \sqrt{\mathrm{Var}(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2 \,\mathrm{Var}(Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2} \\ &\qquad \leq \rho \sqrt{E(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^4 E(Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^4} \leq 32L\rho.\end{aligned}$$

For the remaining pairs $\boldsymbol{s}', \boldsymbol{t}'$ we have a looser bound, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathrm{Cov}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) &\leq \sqrt{\mathrm{Var}(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2 \,\mathrm{Var}(Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2} \\ &\leq 32L.\end{aligned}$$

Recall that P_k is the set of pairs of points $\{\boldsymbol{s}, \boldsymbol{t}\} \in S_k$ with lag in the prescribed range, i.e., $h - \delta_k < |\boldsymbol{s} - \boldsymbol{t}| < h + \delta_k$. We then have the bound

$$\begin{aligned}
\operatorname{Var}(\hat{\gamma}_{\delta_k}(h)) &= \frac{1}{4N(h)^2} \operatorname{Var}\left[\sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} (Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2\right] \\
&= \frac{1}{4N(h)^2} \sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} \sum_{\{\boldsymbol{s}',\boldsymbol{t}'\}\in P_k} \operatorname{Cov}[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2, (Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2] \\
&\leq \frac{1}{4}[(1-\epsilon)(32L\rho) + 32L\epsilon] \\
&= 8L[(1-\epsilon)\rho + \epsilon].
\end{aligned}$$

Letting $\rho \to 0$ and $\epsilon \to 0$ we thus have $\operatorname{Var}(\hat{\gamma}_{\delta_k}) \to 0$.

Now, since γ is continuous (Assumption 3.3), for any $\epsilon > 0$ there is some $\delta_0 > 0$ such that $|\gamma(h+\delta) - \gamma(h)| \leq \epsilon$ for all $\delta \leq \delta_0$. Hence for k sufficiently large so that $\delta_k < \delta_0$, we get

$$\begin{aligned}
E[\hat{\gamma}_{\delta_k}(h)] &= \frac{1}{N(h)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} E(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2 \\
&= \frac{1}{N(h)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} \gamma(|\boldsymbol{s} - \boldsymbol{t}|) \\
&\leq \frac{1}{N(h)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} (\gamma(h) + \epsilon) = \gamma(h) + \epsilon,
\end{aligned}$$

and similarly $E[\hat{\gamma}_{\delta_k}(h)] \geq \gamma(h) - \epsilon$. But as $k \to \infty$ we may take $\epsilon \to 0$, so that $E[\hat{\gamma}_{\delta_k}] \to \gamma(h)$, i.e., $\hat{\gamma}_{\delta_k}(h)$ is asymptotically unbiased. It follows then that

$$E(\hat{\gamma}_{\delta_k}(h) - \gamma(h))^2 = \operatorname{Var}(\hat{\gamma}_{\delta_k}) + [E(\hat{\gamma}_{\delta_k}(h)) - \gamma(h)]^2 \to 0$$

as $k \to \infty$. In other words, $\hat{\gamma}_{\delta k}(h)$ converges to $\gamma(h)$ in L^2, hence in probability. □

Theorem 3.8. *Suppose Assumptions 3.1–3.2 are satisfied. Then there exist sequences C_k and δ_k satisfying the conditions of Theorem 3.1.*

Proof. By Assumption 3.5, choose C_j' such that $\rho^*(C_j') < \frac{1}{j}$. By Assumption 3.1, choose integers k_j sufficiently large so that $p_k(h + C_j' + \frac{1}{j}; h, \frac{1}{j}) < 1/j$ for all $k \geq k_j$, and so that k_j is strictly increasing. Now set $\delta_k = 1/j$ and $C_k = C_j'$, where j is the greatest integer such that $k_j \leq k$. Then $\rho^*(C_k) \to 0, \delta_k \to 0$, and $p_k(h + C_k + \delta_k; h, \delta_k) \to 0$, as desired. □

Theorem 3.2. *If Assumption 3.6 is satisfied and $\delta_k \to 0$, then any sequence $C_k \to \infty$ with $\frac{C_k}{r_k} \to 0$ satisfies the criteria of Theorem 3.1. In particular, if $\delta_k \to 0$, and Assumptions 3.2–3.6 hold, then $\hat{\gamma}_{\delta_k}(h)$ is consistent.*

Proof. The result is trivially true for $h = 0$, so fix $h > 0$. For any ϵ_k-cube E, let $\mathcal{F}_k(E)$ be the collection of ϵ_k-cubes E' such that there exist a pair of points $\boldsymbol{e} \in E$, $\boldsymbol{e}' \in E'$ with $h - \delta_k < d(\boldsymbol{e}, \boldsymbol{e}') < h + \delta_k$. Similarly, let $\mathcal{G}_k(E)$ be the collection of ϵ_k-cubes E' such that for every pair of points $\boldsymbol{e} \in E$, $\boldsymbol{e}' \in E'$, we have $h - \delta_k < d(\boldsymbol{e}, \boldsymbol{e}') < h + \delta_k$.

Claim 1. There are constants b_1, b_2, M_1 such that for all $k \geq M_1$ and every ϵ_k-cube E,

$$b_1 \delta_k / \epsilon_k^n \leq |\mathcal{G}_k(E)| \leq |\mathcal{F}_k(E)| \leq b_2 \delta_k / \epsilon_k^n$$

To prove this claim, fix a k and an ϵ_k-cube E, and consider the set $F = \bigcup_{E' \in \mathcal{F}_k(E)} E'$. Choose a point $\boldsymbol{e} \in E$. For any $\boldsymbol{e}' \in F$, we must have $\boldsymbol{e}' \in E'$ for some $E' \in \mathcal{F}_k(E)$, and by definition this means there exist $\boldsymbol{e}_1 \in E, \boldsymbol{e}_2 \in E'$ with $h - \delta_k < d(\boldsymbol{e}_1, \boldsymbol{e}_2) < h + \delta_k$. Since the cubes E and E' have diameter $\epsilon_k \sqrt{n} \leq \delta_k / 3$ by Assumption 3.6(c), we also have $d(\boldsymbol{e}, \boldsymbol{e}_1) \leq \delta_k / 3$ and $d(\boldsymbol{e}', \boldsymbol{e}_2) \leq \delta_k / 3$, so by the triangle inequality

$$\begin{aligned} d(\boldsymbol{e}, \boldsymbol{e}') &\leq d(\boldsymbol{e}_1, \boldsymbol{e}_2) + d(\boldsymbol{e}, \boldsymbol{e}_1) + d(\boldsymbol{e}', \boldsymbol{e}_2) \\ &< (h + \delta_k) + \delta_k/3 + \delta_k/3 \\ &= h + \frac{5}{3}\delta_k \end{aligned}$$

and likewise,

$$\begin{aligned} d(\boldsymbol{e}, \boldsymbol{e}') &\geq d(\boldsymbol{e}_1, \boldsymbol{e}_2) - d(\boldsymbol{e}, \boldsymbol{e}_1) - d(\boldsymbol{e}', \boldsymbol{e}_2) \\ &> (h - \delta_k) - \delta_k/3 - \delta_k/3 \\ &= h - \frac{5}{3}\delta_k \end{aligned}$$

Therefore,

$$F \subseteq B_{\boldsymbol{e}}(h + \frac{5}{3}\delta_k) \setminus \overline{B}_{\boldsymbol{e}}(h - \frac{5}{3}\delta_k).$$

This implies that $\lambda(F) \leq u(h + \frac{5}{3}\delta_k)^n - u(h - \frac{5}{3}\delta_k)^n$, where u is the measure of the unit ball in $\mathbb{R}^n$. By the mean value theorem, this gives a bound

$$\lambda(F) \leq un \left(h + \frac{5}{3}\delta_k\right)^{n-1} \left(\frac{10}{3}\delta_k\right).$$

On the other hand, $\lambda(F) = \epsilon_k^n |\mathcal{F}_k(E)|$. This gives the desired inequality $|\mathcal{F}_k(E)| \leq b_2 \delta_k / \epsilon_k^n$ for a suitably chosen constant b_2.

Similarly, consider the set $G = \bigcup_{E' \in \mathcal{G}_k(E)} E'$, and choose a point $\boldsymbol{e} \in E$. Choose M_1 so that $\delta_k / 3 < h$ for all $k \geq M_1$. Let k be given with $k \geq M_1$, and also let $\boldsymbol{e}' \in B_{\boldsymbol{e}}\left(h + \frac{1}{3}\delta_k\right) \setminus \overline{B}_{\boldsymbol{e}}\left(h - \frac{1}{3}\delta_k\right)$ be given, so that $h - \frac{1}{3}\delta_k < d(\boldsymbol{e}, \boldsymbol{e}') < h + \frac{1}{3}\delta_k$. Let E' be an ϵ_k-cube containing $\boldsymbol{e}'$. For any $\boldsymbol{e}_1 \in E, \boldsymbol{e}_2 \in E'$ we then have

$$d(\boldsymbol{e}_1, \boldsymbol{e}_2) \leq d(\boldsymbol{e}, \boldsymbol{e}') + d(\boldsymbol{e}, \boldsymbol{e}_1) + d(\boldsymbol{e}', \boldsymbol{e}_2)$$

$$\begin{aligned} &< \left(h + \frac{1}{3}\delta_k\right) + \frac{1}{3}\delta_k + \frac{1}{3}\delta_k \\ &= h + \delta_k, \end{aligned}$$

and

$$\begin{aligned} d(\boldsymbol{e}_1, \boldsymbol{e}_2) &\geq d(\boldsymbol{e}, \boldsymbol{e}') - d(\boldsymbol{e}, \boldsymbol{e}_1) - d(\boldsymbol{e}', \boldsymbol{e}_2) \\ &> \left(h - \frac{1}{3}\delta_k\right) - \frac{1}{3}\delta_k - \frac{1}{3}\delta_k \\ &= h - \delta_k. \end{aligned}$$

Thus $h - \delta_k < d(\boldsymbol{e}_1, \boldsymbol{e}_2) < h + \delta_k$ for all $\boldsymbol{e}_1 \in E, \boldsymbol{e}_2 \in E'$, so by definition $E' \in \mathcal{G}_k(E)$; in particular, $\boldsymbol{e}' \in G$. This proves that

$$G \supseteq B_{\boldsymbol{e}}\left(h + \frac{1}{3}\delta_k\right) \setminus \overline{B}_{\boldsymbol{e}}\left(h - \frac{1}{3}\delta_k\right).$$

Therefore, $\lambda(G) \geq u(h + \frac{1}{3}\delta_k)^n - u(h - \frac{1}{3}\delta_k)^n$. The mean value theorem gives a bound

$$\lambda(G) \geq un\left(h - \frac{1}{3}\delta_k\right)^{n-1}\left(\frac{2}{3}\delta_k\right)$$

On the other hand, $\lambda(G) = \epsilon_k^n |\mathcal{G}_k(E)|$. This gives the desired inequality $|\mathcal{G}_k(E)| \geq b_1\delta_k/\epsilon_k^n$ for all $k \geq M_1$, for a suitably chosen constant b_1, proving the claim.

Now let $\mathcal{F}'_k$ be the collection of ϵ_k-cubes E intersecting V_k, and let $\mathcal{G}'_k$ be the collection of ϵ_k-cubes E such that E together with all the cubes in $\mathcal{G}_k(E)$ are contained in V_k.

Claim 2. There are positive constants a_1, a_2, M_2 such that for all $i \geq M_2$,

$$a_1\lambda(V_k)/\epsilon_k^n \leq |\mathcal{G}'_k| \leq |\mathcal{F}'_k| \leq a_2\lambda(V_k)/\epsilon_k^n$$

To prove this claim, fix k and let $F = \bigcup_{E\in\mathcal{F}'_k} E$. Then, referring to Assumption 3.6(b), we have $F \subseteq B_{\boldsymbol{x}_k}(r_kK + \delta_k/3)$, hence $\lambda(F) \leq u(r_kK + \delta_k/3)^n$. Also $V_k \supseteq B_{\boldsymbol{x}_k}(r_k/K)$, hence $\lambda(V_k) \geq u(r_k/K)^n$. On the other hand, $\lambda(F) = \epsilon_k^n|\mathcal{F}'_k|$. Together this gives

$$|\mathcal{F}'_k| \leq u(r_kK + \delta_k/3)^n/\epsilon_k^n \leq u(r_kK + \delta_k/3)^n \frac{\lambda(V_k)}{u(r_k/K)^n}/\epsilon_k^n$$

Since $(r_kK + \delta_k/3)^n/(r_k/K)^n \to K^{2n}$ as $k \to \infty$, this gives the desired bound $|\mathcal{F}'_k| \leq a_2\lambda(V_k)/\epsilon_k^n$ for some constant a_2.

Similarly, let $G = \bigcup_{E\in\mathcal{G}'_k} E$. For sufficiently large k, we have $r_k/K - h - 2\delta_k/3 > 0$, in which case $G \supseteq B_{\boldsymbol{x}_k}(r_k/K - h - 2\delta_k/3)$, hence $\lambda(G) \geq u(r_k/K - h - 2\delta_k/3)^n$. Also $V_k \subseteq B_{\boldsymbol{x}_k}(r_kK)$, hence $\lambda(V_k) \leq u(r_kK)^n$. On the other hand, $\lambda(G) = \epsilon_k^n|\mathcal{G}'_k|$. Together this gives

$$|\mathcal{G}'_k| \geq u(r_k/K - h - 2\delta_k/3)^n/\epsilon_k^n$$

$$\geq u(r_k/K - h - 2\delta_k/3)^n \frac{\lambda(V_k)}{u(r_k K)^n}/\epsilon_k^n$$

Since $(r_k/K - h - 2\delta_k/3)/(r_k K)^n \to K^{-2n}$ as $k \to \infty$, this gives the desired bound $|\mathcal{G}'_k| \geq a_1\lambda(V_k)/\epsilon_k^n$ for some constant a_1, for all sufficiently large k (say, $k \geq M_2$). This proves the claim.

Now given an $\boldsymbol{x} \in \mathbb{R}^n$, let $\mathcal{H}_k(\boldsymbol{x})$ be the collection of closed ϵ_k-cubes E such that there exists a point $\boldsymbol{e} \in E$ with $d(\boldsymbol{x}, \boldsymbol{e}) < h + C_k + \delta_k$.

Claim 3. There is a positive constant a_3 such that for all k and all $\boldsymbol{x} \in \mathbb{R}^n$,

$$|\mathcal{H}_k(\boldsymbol{x})| \leq a_3(h + C_k)^n/\epsilon_k^n$$

Let $H = \bigcup_{E \in \mathcal{H}_k(\boldsymbol{x})} E$. Then

$$H \subseteq B_{\boldsymbol{x}}(h + C_k + \delta_k + \delta_k/3)$$

Thus $\lambda(H) \leq u(h + C_k + 4\delta_k/3)^n$. On the other hand, $\lambda(H) = \epsilon_k^n |\mathcal{H}_k(\boldsymbol{x})|$. Therefore, $|\mathcal{H}_k(\boldsymbol{x})| \leq u(h + C_k + 4\delta_k/3)^n/\epsilon_k^n$, and the claim follows, since $\delta_k \to 0$.

Now, by Assumption 3.6(c), for all $k \geq M$ the number of points of S_k in each ϵ_k-cube $E \in \mathcal{G}'_k$ is at least $A_1|S_k|\epsilon_k^n/\lambda(V_k)$, and the same lower bound holds for each cube in $\mathcal{G}_k(E)$. Therefore, we obtain the following bound for the number of unordered pairs of points $\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq S_k$ with $h - \delta_k < d(\boldsymbol{s}, \boldsymbol{t}) < h + \delta_k$:

$$\begin{aligned}
|P_k| &\geq \frac{1}{2} \sum_{E \in \mathcal{G}'_k} \sum_{E' \in \mathcal{G}_k(E)} |E \cap S_k||E' \cap S_k| \\
&\geq \frac{1}{2} \sum_{E \in \mathcal{G}'_k} |\mathcal{G}_k(E)| \left(A_1|S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2 \\
&\geq \frac{1}{2} |\mathcal{G}'_k| \frac{b_1\delta_k}{\epsilon_k^n} \left(A_1|S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2 \\
&\geq \frac{1}{2} \frac{a_1\lambda(V_k)}{\epsilon_k^n} \frac{b_1\delta_k}{\epsilon_k^n} \left(A_1|S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2
\end{aligned}$$

However, given a ball U of radius $h + C_k + \delta_k$, centered at $\boldsymbol{x} \in \mathbb{R}^n$, the number of closed ϵ_k-cubes intersecting U is $|\mathcal{H}_k(\boldsymbol{x})| \leq a_3(h + C_k)^n/\epsilon_k^n$, and each such cube contains at most $A_2|S_k|\epsilon_k^n/\lambda(V_k)$ points of S_k. Therefore, the number of pairs $\{\boldsymbol{s}, \boldsymbol{t}\} \in P_k$ such that either $\boldsymbol{s} \in U$ or $\boldsymbol{t} \in U$ is no more than

$$\sum_{E \in \mathcal{H}_k(\boldsymbol{x})} \sum_{E' \in \mathcal{F}_k(E)} |E \cap S_k||E' \cap S_k|$$

$$\leq \sum_{E \in \mathcal{H}_k(\boldsymbol{x})} |\mathcal{F}_k(E)| \left(A_2 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2$$
$$\leq |\mathcal{H}_k(\boldsymbol{x})| \frac{b_2 \delta_k}{\epsilon_k^n} \left(A_2 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2$$
$$\leq \frac{a_3 (h + C_k)^n}{\epsilon_k^n} \frac{b_2 \delta_k}{\epsilon_k^n} \left(A_2 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2$$

Therefore, the proportion satisfies the bound

$$p_k(h + C_k + \delta_k; h, \delta_k) \leq \frac{\frac{a_3 (h + C_k)^n}{\epsilon_k^n} \frac{b_2 \delta_k}{\epsilon_k^n} \left(A_2 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2}{\frac{1}{2} \frac{a_1 \lambda(V_k)}{\epsilon_k^n} \frac{b_1 \delta_k}{\epsilon_k^n} \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)} \right)^2}$$
$$\asymp \frac{C_k^n}{\lambda(V_k)} \asymp \left(\frac{C_k}{r_k} \right)^n$$

which approaches 0 as $k \to \infty$, by assumption. □

For the proof of Theorem 3.3, we will need the following Chernoff bound [7]:

Lemma 3.9. *For any $\alpha > 0$, there is a constant $v < 1$ such that for any binomial random variable X,*

$$P(|X - E(X)| \geq \alpha E(X)) \leq 2v^{E(X)}$$

Equivalently, there is a constant $C > 0$ such that

$$P(|X - E(X)| \geq \alpha E(X)) \leq 2e^{-CE(X)}$$

Proof. Without loss of generality, we can express X as a sum $X = \sum_{i=1}^m X_i$ of i.i.d. Bernoulli random variables $X_1, \ldots, X_m$ with some parameter $0 \leq p \leq 1$. Write $\mu = E(X) = mp$. Then

$$E((1+\alpha)^X) = E((1+\alpha)^{\sum_{i=1}^m X_i})$$
$$= \prod_{i=1}^m E((1+\alpha)^{X_i})$$
$$= (1 - p + p(1+\alpha))^m$$
$$= (1 + p\alpha)^m \leq e^{p\alpha m} = e^{\mu\alpha}$$

Therefore, by Markov's inequality,

$$P(X \geq (1+\alpha)\mu) = P\left((1+\alpha)^X \geq (1+\alpha)^{(1+\alpha)\mu} \right)$$
$$\leq \frac{E((1+\alpha)^X)}{(1+\alpha)^{(1+\alpha)\mu}}$$

$$\leq \frac{e^{\alpha\mu}}{(1+\alpha)^{(1+\alpha)\mu}}$$
$$= \left(\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right)^{\mu}$$

Claim. For any $y \neq 0$, we have $\frac{e^y}{(1+y)^{1+y}} < 1$.

If we define

$$f(y) = \log\left(\frac{e^y}{(1+y)^{1+y}}\right) = y - (1+y)\log(1+y)$$

then we see that $f(0) = 0$, while

$$f'(y) = -\log(1+y)$$

which is positive for $y < 0$ and negative for $y > 0$, so that $f(y) < 0$ for all $y \neq 0$, proving the claim.

Without loss of generality, we may assume $\alpha < 1$. Now apply Markov's inequality again to obtain

$$P(X \leq (1-\alpha)\mu) = P\left((1-\alpha)^X \geq (1-\alpha)^{(1-\alpha)\mu}\right)$$
$$\leq \frac{E((1-\alpha)^X)}{(1-\alpha)^{(1-\alpha)\mu}}$$
$$\leq \frac{e^{-\alpha\mu}}{(1-\alpha)^{(1-\alpha)\mu}}$$
$$= \left(\frac{e^{-\alpha}}{(1-\alpha)^{1-\alpha}}\right)^{\mu}.$$

By the claim, we have both $\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}} < 1$ and $\frac{e^{-\alpha}}{(1-\alpha)^{1-\alpha}} < 1$, so if we set v to be their maximum, then we have

$$P(|X - E(X)| \geq \alpha E(X)) = P(X \geq (1+\alpha)\mu) + P(X \leq (1-\alpha)\mu)$$
$$\leq \left(\frac{e^{\alpha}}{(1+\alpha)^{1+\alpha}}\right)^{\mu} + \left(\frac{e^{-\alpha}}{(1-\alpha)^{1-\alpha}}\right)^{\mu} \leq 2v^{\mu}$$

□

Lemma 3.10. *Let $X \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathrm{Bin}(n,q)$ be binomial random variables with $p \leq q$. For any $t \in \mathbb{R}$, we have $P(X \geq t) \leq P(Y \geq t)$.*

Proof. Let $Z = (Z_1, Z_2, Z_3)$ be a multinomial random variable with parameters n and $(p, q-p, 1-q)$. Then $X \overset{d}{=} Z_1$ and $Y \overset{d}{=} Z_1 + Z_2$, and $\{Z_1 \geq t\} \subseteq \{Z_1 + Z_2 \geq t\}$, so that

$$P(X \geq t) = P(Z_1 \geq t) \leq P(Z_1 + Z_2 \geq t) = P(Y \geq t).$$

□

Theorem 3.3. *Suppose there are constants $0 < A_1' \le A_2' < \infty$ and measurable functions $f_k : V_k \to \mathbb{R}$, with $A_1' \le f_k \le A_2'$, and suppose Assumptions 3.6(a,b) and 3.7 are satisfied. Suppose S_1, S_2, ..., are random sets where S_k is distributed according to an inhomogeneous binomial point process with intensity $f_k(x)/\int_{V_k} f_k \, d\lambda$ and $|S_k| = m_k$. Then with probability 1, Assumption 3.6(c) is also satisfied for any choice of sequence δ_k satisfying $\epsilon_k \le \frac{\delta_k}{3\sqrt{n}}$.*

Proof. Let $L = A_2'/A_1'$. Given an ϵ_k-cube $G \subseteq V_k$, the number of points $|S_k \cap G|$ is binomial random variable with

$$E(|S_k \cap G|) = \frac{m_k}{\lambda(V_k)} \int_G f_k \, d\lambda \in \frac{m_k \epsilon_k^n}{\lambda(V_k)} \left[\frac{1}{L}, L\right].$$

Set $A_1 = \frac{1}{2L}$ and $A_2 = \frac{3}{2}L$. Then by Lemma 3.9,

$$\begin{aligned}
&P\left(|S_k \cap G| \notin m_k \epsilon_k^n [A_1, A_2]/\lambda(V_k)\right) \\
&\le P\left(\Big||S_k \cap G| - E(|S_k \cap G|)\Big| \ge \frac{1}{2} E(|S_k \cap G|)\right) \\
&\le 2\exp\left(-CE(|S_k \cap G|)\right) \\
&\le 2\exp\left(-\frac{Cm_k\epsilon_k^n/L}{\lambda(V_k)}\right). \qquad (3.4)
\end{aligned}$$

for some constant $C > 0$. Now, let G be any ϵ_k-cube intersecting V_k (but not necessarily contained in V_k), the number of points $|S_k \cap G|$ again has a binomial distribution $\mathrm{Bin}(m_k, p)$ for some $p \le \epsilon_k^n L/\lambda(V_k)$. For sufficiently large k, we have $\epsilon_k^n L/\lambda(V_k) \le 1$; in this case, let Y be a random variable with distribution $\mathrm{Bin}(m_k, q)$, where $q = \epsilon_k^n L/\lambda(V_k)$, so that Lemma 3.10 implies

$$\begin{aligned}
&P(|S_k \cap G| > m_k \epsilon_k^n A_2/\lambda(V_k)) \\
&\le P(Y > m_k \epsilon_k^n A_2/\lambda(V_k)) \\
&= P(Y > \frac{3}{2}E(Y)) \\
&\le P(|Y - E(Y)| \le \frac{1}{2}E(Y)) \\
&\le 2\exp(-CE(Y)) \\
&\le 2\exp\left(-\frac{Cm_k\epsilon_k^n L}{\lambda(V_k)}\right) \\
&\le 2\exp\left(-\frac{Cm_k\epsilon_k^n/L}{\lambda(V_k)}\right)
\end{aligned}$$

By Assumption 3.7, for sufficiently large k we have

$$\frac{m_k\epsilon_k^n/\lambda(V_k)}{\log m_k} \ge \frac{3L}{C}$$

So, given k and an ϵ_k-cube G, the probability that it violates the criterion in Assumption 3.6(c) is no more than

$$\begin{aligned} & 2\exp\left(-\frac{Cm_k\epsilon_k^n/L}{\lambda(V_k)}\right) \\ & \leq 2\exp\left(-3\log m_k\right) \\ & = 2/m_k^3 \end{aligned}$$

As we showed in Claim 2 of the proof of Theorem 3.2, there is a constant a_2 such that the number of ϵ_k-cubes intersecting V_k is no more than $a_2\lambda(V_k)/\epsilon_k^n$. Therefore, for any fixed sufficiently large k, by Boole's inequality, the probability that the criterion of Assumption 3.6(c) fails (for at least one ϵ_k-cube) for this k is no more than

$$\begin{aligned} & \frac{a_2\lambda(V_k)}{\epsilon_k^n}\frac{2}{m_k^3} \\ & \leq \frac{a_2 m_k C}{3L\log m_k}\frac{2}{m_k^3} \\ & = \frac{2a_2 C}{3Lm_k^2\log m_k} \end{aligned}$$

Summing this over all k gives a finite sum, so by the Borel-Cantelli Lemma, Assumption 3.6(c) is satisfied with probability 1. □

To prepare for the proof of Theorem 3.4, we will need estimates for the number of pairs of points separated by a distance in a given interval. Given location sets $S_1, S_2, \ldots$ and a subset $H \subseteq \mathbb{R}$, let $N_k(H)$ denote the number of unordered pairs of points $\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq S_k$ with $d(\boldsymbol{s}, \boldsymbol{t}) \in H$.

Lemma 3.11. *Suppose that Assumption 3.6 is satisfied and $\delta_k \to 0$, and fix an $h_0 > 0$. Then there are constants $D_1, D_2, M_1 > 0$ such that for all $h_1, h_2 \in [0, h_0]$ with $h_1 < h_2$, if $k \geq M_1$ then*

$$N_k([h_1,h_2]) \geq D_1\frac{|S_k|^2}{\lambda(V_k)}\lambda\left(B_{\mathbf{0}}(h_2-2\epsilon_k\sqrt{n})\setminus B_{\mathbf{0}}(h_1+2\epsilon_k\sqrt{n})\right)$$

$$N_k([h_1,h_2]) \leq D_2\frac{|S_k|^2}{\lambda(V_k)}\lambda\left(B_{\mathbf{0}}(h_2+2\epsilon_k\sqrt{n})\setminus B_{\mathbf{0}}(h_1-2\epsilon_k\sqrt{n})\right)$$

In addition, if $h_2 - h_1 \geq 4\epsilon_k\sqrt{n}$, then

$$N_k([h_1,h_2]) \geq D_1\frac{|S_k|^2}{\lambda(V_k)}nu\left(h_1+2\epsilon_k\sqrt{n}\right)^{n-1}\left(h_2-h_1-4\epsilon_k\sqrt{n}\right)$$

$$N_k([h_1,h_2]) \leq D_2\frac{|S_k|^2}{\lambda(V_k)}nu\left(h_2+2\epsilon_k\sqrt{n}\right)^{n-1}\left(h_2-h_1+4\epsilon_k\sqrt{n}\right)$$

where u is the measure of the unit ball in $\mathbb{R}^n$. (Here we define $B_{\mathbf{0}}(r)$ to be the empty set if $r \leq 0$.)

Remark: These bounds are continuous functions of h_1 and h_2; therefore, if we were to replace the closed interval $[h_1, h_2]$ by an open interval or a half-open interval, the same bounds would apply.

Proof. For any ϵ_k-cube E, let $\mathcal{F}_k(E) = \mathcal{F}_k(E, h_1, h_2)$ be the collection of ϵ_k-cubes E' such that there exist a pair of points $e \in E$, $e' \in E'$ with $h_1 < d(e, e') < h_2$. Similarly, let $\mathcal{G}_k(E) = \mathcal{G}_k(E, h_1, h_2)$ be the collection of ϵ_k-cubes E' such that for every pair of points $e \in E$, $e' \in E'$, we have $h_1 < d(e, e') < h_2$.

Claim 1. If $h_2 - h_1 \geq 4\epsilon_k\sqrt{n}$, then for any ϵ_k-cube E,

$$|\mathcal{F}_k(E)| \leq \lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)\epsilon_k^{-n}$$
$$|\mathcal{G}_k(E)| \geq \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right)\epsilon_k^{-n}$$

To prove this claim, fix k and consider the set $F = \bigcup_{E' \in \mathcal{F}_k(E)} E'$. Fix a point $\boldsymbol{e} \in E$. Since each ϵ_k cube has diameter $\epsilon_k\sqrt{n}$, we have

$$F \subseteq B_{\boldsymbol{e}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\boldsymbol{e}}(h_1 - 2\epsilon_k\sqrt{n}).$$

Therefore, $\lambda(F) \leq \lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)$. On the other hand, $\lambda(F) = \epsilon_k^n|\mathcal{F}_k(E)|$. This gives the desired inequality for $|\mathcal{F}_k(E)|$.

Similarly, consider the set $G = \bigcup_{E' \in \mathcal{G}_k(E)} E'$, and fix a point $\boldsymbol{e} \in E$. We have

$$G \supseteq B_{\boldsymbol{e}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\boldsymbol{e}}(h_1 + 2\epsilon_k\sqrt{n})$$

Therefore, $\lambda(G) \geq \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right)$. On the other hand, $\lambda(G) = \epsilon_k^n|\mathcal{G}_k(E)|$. This gives the desired inequality for $|\mathcal{G}_k(E)|$, proving the claim.

As in the proof of Theorem 3.2, let $\mathcal{F}'_k$ be the collection of ϵ_k-cubes intersecting V_k, and let $\mathcal{G}'_k = \mathcal{G}'_k(h_1, h_2)$ be the collection of open ϵ_k-cubes E such that E together with all the cubes in $\mathcal{G}_k(E)$ are contained in V_k.

Claim 2. There are positive constants a_1, a_2, M_1 (not depending on h_1 and h_2) such that for all $k \geq M_1$,

$$a_1\lambda(V_k)/\epsilon_k^n \leq |\mathcal{G}'_k| \leq |\mathcal{F}'_k| \leq a_2\lambda(V_k)/\epsilon_k^n$$

To prove this claim, fix k and let $F = \bigcup_{E\in\mathcal{F}'_k} E$. Then, referring to Assumption 3.6(b), we have $F \subseteq B_{\boldsymbol{x}_k}(r_kK + \epsilon_k\sqrt{n})$, hence $\lambda(F) \le u(r_kK + \epsilon_k\sqrt{n})^n$. Also $V_k \supseteq B_{\boldsymbol{x}_k}(r_k/K)$, hence $\lambda(V_k) \ge u(r_k/K)^n$. On the other hand, $\lambda(F) = \epsilon_k^n|\mathcal{F}'_k|$. Together this gives

$$|\mathcal{F}'_k| \le u(r_kK + \epsilon_k\sqrt{n})^n/\epsilon_k^n \le u(r_kK + \epsilon_k\sqrt{n})^n \frac{\lambda(V_k)}{u(r_k/K)^n}/\epsilon_k^n$$

Since $(r_kK + \epsilon_k\sqrt{n})^n/(r_k/K)^n \to K^{2n}$ as $k \to \infty$, this gives the desired bound $|\mathcal{F}'_k| \le a_2\lambda(V_k)/\epsilon_k^n$, for a suitable constant a_2.

Similarly, let $G = G(h_1, h_2)$ by $G = \bigcup_{E\in\mathcal{G}'_k} E$. Then $G \supseteq B_{\boldsymbol{x}_k}(r_k/K - h_2 - 2\epsilon_k\sqrt{n})$, hence $\lambda(G) \ge u(r_k/K - h_2 - 2\epsilon_k\sqrt{n})^n$. Also $V_k \subseteq B_{\boldsymbol{x}_k}(r_kK)$, hence $\lambda(V_k) \le u(r_kK)^n$. On the other hand, $\lambda(G) = \epsilon_k^n|\mathcal{G}'_k|$. Together this gives

$$\begin{aligned}|\mathcal{G}'_k| &\ge u(r_k/K - h_2 - 2\epsilon_k\sqrt{n})^n/\epsilon_k^n \\ &\ge u(r_k/K - h_2 - 2\epsilon_k\sqrt{n})^n \frac{\lambda(V_k)}{u(r_kK)^n}/\epsilon_k^n\end{aligned}$$

Since $\bigcup_{k=1}^{\infty} V_k = \mathbb{R}^n$ by Assumption 3.6(a), it follows from Assumption 3.6(b) that $r_k \to \infty$ as $k \to \infty$, so that for sufficiently large k, we have $r_k/K - h_0 - 2\epsilon_k\sqrt{n} > 0$, in which case

$$|\mathcal{G}'_k| \ge u(r_k/K - h_0 - 2\epsilon_k\sqrt{n})^n \frac{\lambda(V_k)}{u(r_kK)^n}/\epsilon_k^n$$

Since $(r_k/K - h_0 - 2\epsilon_k\sqrt{n})^n/(r_kK)^n \to K^{-2n}$ as $k \to \infty$, this gives the desired bound $|\mathcal{G}'_k| \ge a_1\lambda(V_k)/\epsilon_k^n$ for $k \ge M_1$, for suitable constants M_1 and a_1. This proves the claim.

Without loss of generality, we may assume that M_1 has been chosen sufficiently large that $M_1 \ge M$, where M is the constant from Assumption 3.6(c). Now we have that for all $k \ge M_1$,

$$\begin{aligned}N_k([h_1, h_2]) &\le \frac{1}{2}\sum_{E\in\mathcal{F}'_k}\sum_{E'\in\mathcal{F}_k(E)} |E \cap S_k||E' \cap S_k| \\ &\le \frac{1}{2}\sum_{E\in\mathcal{F}'_k}\sum_{E'\in\mathcal{F}_k(E)} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &= \frac{1}{2}\sum_{E\in\mathcal{F}'_k} |F_k(E)| \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &\le \frac{1}{2}\sum_{E\in\mathcal{F}'_k} \lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)\epsilon_k^{-n}\left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &= \frac{1}{2}|\mathcal{F}'_k|\lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)\epsilon_k^{-n}\left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &\le \frac{1}{2}a_2\lambda(V_k)\epsilon_k^{-n}\lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)\epsilon_k^{-n}\left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2\end{aligned}$$

$$= \frac{1}{2} a_2 A_2^2 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h_2 + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 - 2\epsilon_k\sqrt{n})\right)$$

which proves the desired upper bound on $N_k([h_1, h_2])$, setting $D_2 = \frac{1}{2} a_2 A_2^2$. Similarly,

$$\begin{aligned}
N_k([h_1, h_2]) &\geq \frac{1}{2} \sum_{E \in \mathcal{G}_k'} \sum_{E' \in \mathcal{G}_k(E)} |E \cap S_k||E' \cap S_k| \\
&\geq \frac{1}{2} \sum_{E \in \mathcal{G}_k'} \sum_{E' \in \mathcal{G}_k(E)} \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\
&= \frac{1}{2} \sum_{E \in \mathcal{G}_k'} |\mathcal{G}_k(E)| \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\
&\geq \frac{1}{2} \sum_{E \in \mathcal{G}_k'} \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\
&= \frac{1}{2} |\mathcal{G}_k'| \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\
&\geq \frac{1}{2} a_1 \lambda(V_k) \epsilon_k^{-n} \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_1 |S_k| \frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\
&= \frac{1}{2} a_1 A_1^2 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h_2 - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h_1 + 2\epsilon_k\sqrt{n})\right)
\end{aligned}$$

which proves the desired lower bound on $N_k([h_1, h_2])$, setting $D_1 = \frac{1}{2} a_1 A_1^2$. The latter part of the theorem follows by the mean value theorem. □

Now we will need to adapt the proof of Theorem 3.2 to obtain an explicit bound on the proportions $p_k(r; h, \delta_k)$:

Theorem 3.12. *Suppose that Assumption 3.6 is satisfied and $\delta_k \to 0$. For a suitably chosen positive constant D_3 (not depending on r, h, or k), the following inequality holds:*

$$p_k(r; h, \delta_k) \leq D_3 \frac{(r + 2\epsilon_k\sqrt{n})^n}{\lambda(V_k)}$$

Proof. Given an $\boldsymbol{x} \in \mathbb{R}^n$, let $\mathcal{H}_k(\boldsymbol{x}) = \mathcal{H}_k(\boldsymbol{x}, r)$ be the collection of ϵ_k-cubes E such that there exists a point $\boldsymbol{e} \in E$ with $d(\boldsymbol{x}, \boldsymbol{e}) < r$. We claim that

$$|\mathcal{H}_k(\boldsymbol{x})| \leq u(r + \epsilon_k\sqrt{n})^n / \epsilon_k^n$$

To see this, let $H = \bigcup_{E \in \mathcal{H}_k(\boldsymbol{x})} E$; then

$$H \subseteq B_{\boldsymbol{x}}(r + \epsilon_k\sqrt{n}),$$

and hence $\lambda(H) \leq u(r + \epsilon_k\sqrt{n})^n$. On the other hand, $\lambda(H) = \epsilon_k^n |\mathcal{H}_k(\boldsymbol{x})|$. Therefore, $|\mathcal{H}_k(\boldsymbol{x})| \leq u(r + \epsilon_k\sqrt{n})^n / \epsilon_k^n$, which proves the claim.

Consider first the case in which $h - \delta_k + 2\epsilon_k\sqrt{n} \geq 0$. Applying Lemma 3.11 with $h_1 = h - \delta_k$ and $h_2 = h + \delta_k$, and using the elementary inequality $b^n - a^n \geq b^{n-1}(b-a)$ which holds for all $0 \leq a \leq b$, we have

$$\begin{aligned}|P_k| = N_k((h_1, h_2)) &\geq D_1 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h + \delta_k - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h - \delta_k + 2\epsilon_k\sqrt{n})\right) \\ &\geq D_1 \frac{|S_k|^2}{\lambda(V_k)} u(h + \delta_k - 2\epsilon_k\sqrt{n})^{n-1}(2\delta_k - 4\epsilon_k\sqrt{n})\end{aligned}$$

Given a ball U of radius r, centered at $\boldsymbol{x} \in \mathbb{R}^n$, the number of ϵ_k-cubes intersecting U is $|\mathcal{H}_k(\boldsymbol{x})| \leq u(r + 2\epsilon_k\sqrt{n})^n/\epsilon_k^n$, and each such cube contains at most $A_2|S_k|\epsilon_k^n/\lambda(V_k)$ points of S_k. Therefore, using Claim 1 from the proof of Lemma 3.11, the number of unordered pairs of distinct points $\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq S_k$ with $h - \delta_k < |\boldsymbol{s} - \boldsymbol{t}| < h + \delta_k$ and either $\boldsymbol{s} \in U$ or $\boldsymbol{t} \in U$ is no more than

$$\begin{aligned}&\sum_{E \in \mathcal{H}_k(\boldsymbol{x})} \sum_{E' \in \mathcal{F}_k(E)} |E \cap S_k||E' \cap S_k| \\ &\leq \sum_{E \in \mathcal{H}_k(\boldsymbol{x})} |\mathcal{F}_k(E)| \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &\leq |\mathcal{H}_k(x)|\lambda\left(B_{\mathbf{0}}(h + \delta_k + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h - \delta_k - 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &\leq |\mathcal{H}_k(x)|un\left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1}\left(2\delta_k + 4\epsilon_k\sqrt{n}\right)\epsilon_k^{-n}\left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &\leq \frac{u(r + \epsilon_k\sqrt{n})^n}{\epsilon_k^n} un\left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1}\left(2\delta_k + 4\epsilon_k\sqrt{n}\right)\epsilon_k^{-n}\left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2 \\ &= \frac{u^2 n A_2^2 |S_k|^2}{(\lambda(V_k))^2}(r + \epsilon_k\sqrt{n})^n\left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1}\left(2\delta_k + 4\epsilon_k\sqrt{n}\right)\end{aligned}$$

Therefore, the proportion satisfies the bound

$$\begin{aligned}p_k(r; h, \delta_k) &\leq \frac{\frac{u^2 n A_2^2 |S_k|^2}{(\lambda(V_k))^2}(r + \epsilon_k\sqrt{n})^n\left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1}\left(2\delta_k + 4\epsilon_k\sqrt{n}\right)}{D_1\frac{|S_k|^2}{\lambda(V_k)} u\left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1}\left(2\delta_k - 4\epsilon_k\sqrt{n}\right)} \\ &\leq D_3\frac{(r + \epsilon_k\sqrt{n})^n}{\lambda(V_k)}\end{aligned}$$

for a suitably chosen constant $D_3 > 0$, where we used the fact that $\epsilon_k \leq \frac{\delta_k}{3\sqrt{n}}$ (from Assumption 3.6(c)).

Now consider the case in which $h - \delta_k + 2\epsilon_k\sqrt{n} < 0$. Applying Lemma 3.11, we have

$$|P_k| = N_k((h_1, h_2)) \geq D_1\frac{|S_k|^2}{\lambda(V_k)}\lambda\left(B_{\mathbf{0}}(h + \delta_k - 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h - \delta_k + 2\epsilon_k\sqrt{n})\right)$$

$$= D_1 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h + \delta_k - 2\epsilon_k\sqrt{n})\right)$$
$$= D_1 \frac{|S_k|^2}{\lambda(V_k)} u(h + \delta_k - 2\epsilon_k\sqrt{n})^n$$

On the other hand, the number of unordered pairs of distinct points $\{\boldsymbol{s}, \boldsymbol{t}\} \subseteq S_k$ with $h - \delta_k < |\boldsymbol{s} - \boldsymbol{t}| < h + \delta_k$ and either $\boldsymbol{s} \in U$ or $\boldsymbol{t} \in U$ is no more than

$$\sum_{E \in \mathcal{H}_k(\boldsymbol{x})} \sum_{E' \in \mathcal{F}_k(E)} |E \cap S_k||E' \cap S_k|$$
$$\leq \sum_{E \in \mathcal{H}_k(\boldsymbol{x})} |\mathcal{F}_k(E)| \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2$$
$$\leq |\mathcal{H}_k(x)|\lambda\left(B_{\mathbf{0}}(h + \delta_k + 2\epsilon_k\sqrt{n}) \setminus B_{\mathbf{0}}(h - \delta_k - 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2$$
$$\leq |\mathcal{H}_k(x)|\lambda\left(B_{\mathbf{0}}(h + \delta_k + 2\epsilon_k\sqrt{n})\right) \epsilon_k^{-n} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2$$
$$= |\mathcal{H}_k(x)|u(h + \delta_k + 2\epsilon_k\sqrt{n})^n \epsilon_k^{-n} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2$$
$$\leq \frac{u(r + \epsilon_k\sqrt{n})^n}{\epsilon_k^n} u(h + \delta_k + 2\epsilon_k\sqrt{n})^n \epsilon_k^{-n} \left(A_2|S_k|\frac{\epsilon_k^n}{\lambda(V_k)}\right)^2$$
$$= \frac{u^2 A_2^2 |S_k|^2}{(\lambda(V_k))^2} (r + \epsilon_k\sqrt{n})^n \left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^n$$

Therefore, the proportion satisfies the bound

$$p_k(r; h, \delta_k) \leq \frac{\frac{u^2 A_2^2 |S_k|^2}{(\lambda(V_k))^2} (r + \epsilon_k\sqrt{n})^n \left(h + \delta_k + 2\epsilon_k\sqrt{n}\right)^n}{D_1 \frac{|S_k|^2}{\lambda(V_k)} u(h + \delta_k - 2\epsilon_k\sqrt{n})^n}$$
$$\leq D_3' \frac{(r + \epsilon_k\sqrt{n})^n}{\lambda(V_k)}$$

for a suitably chosen constant $D_3' > 0$. Replacing D_3 by the maximum of D_3 and D_3', the proof is complete. □

Now we need to adapt the proof of Theorem 3.1 to obtain an explicit bound on the variance of $\hat{\gamma}_\delta(h)$:

Theorem 3.13. *Suppose that Assumption 3.6 is satisfied and $\delta_k \to 0$, and fix a number $h_0 > 0$. Given a sequence $C_k \to \infty$, there is a suitable constant D_4 such that for all $h \in [0, h_0]$ and all k,*

$$\mathrm{Var}(\hat{\gamma}_\delta(h)) \leq D_4 \left(\rho^*(C_k) + \left(\frac{C_k}{r_k}\right)^n\right)$$

Proof. Let $L = E(Z(s)^4) < \infty$. We note that for any $\boldsymbol{s}, \boldsymbol{t}$, $E[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^4] \leq 32L$. Let $\epsilon = p_k(h + \delta_k + C_k; h, \delta_k)$. Given a pair of points $\boldsymbol{s}, \boldsymbol{t} \in S_k$ with $h - \delta_k < \|\boldsymbol{s} - \boldsymbol{t}\| < h + \delta_k$, taking $U = B_{\boldsymbol{s}}(h + \delta_k + C_k)$ we know that at least proportion $1 - \epsilon$ of the pairs $\boldsymbol{s}', \boldsymbol{t}'$ with $h - \delta < |\boldsymbol{s}' - \boldsymbol{t}'| < h + \delta$ have $d(\{\boldsymbol{s}, \boldsymbol{t}\}, \{\boldsymbol{s}', \boldsymbol{t}'\}) \geq C_k$, and for these, we may bound the covariance

$$\begin{aligned}
&\mathrm{Cov}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) \\
&\qquad = \mathrm{Corr}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) \\
&\qquad\quad \cdot \sqrt{\mathrm{Var}[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2]\,\mathrm{Var}[(Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2]} \\
&\qquad \leq \rho^*(C_k)\sqrt{E[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^4]E[(Z(\boldsymbol{s}') - Z(\boldsymbol{t}')^4]} \\
&\qquad \leq 32L\rho^*(C_k).
\end{aligned}$$

For the remaining pairs $\boldsymbol{s}', \boldsymbol{t}'$ we have a looser bound, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathrm{Cov}([Z(\boldsymbol{s}) - Z(\boldsymbol{t})]^2, [Z(\boldsymbol{s}') - Z(\boldsymbol{t}')]^2) &\leq \sqrt{\mathrm{Var}[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2]\,\mathrm{Var}[(Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2]} \\
&\leq 32L.
\end{aligned}$$

Recall that P_k is the set of pairs of points $\{\boldsymbol{s}, \boldsymbol{t}\} \in S_k$ with lag in the prescribed range, i.e., $h - \delta_k < \|\boldsymbol{s} - \boldsymbol{t}\| < h + \delta_k$. We then have the bound

$$\begin{aligned}
\mathrm{Var}(\hat{\gamma}_{\delta_k}(h)) &= \frac{1}{4N(h)^2}\,\mathrm{Var}\left[\sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k} (Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2\right] \\
&= \frac{1}{4N(h)^2}\sum_{\{\boldsymbol{s},\boldsymbol{t}\}\in P_k}\sum_{\{\boldsymbol{s}',\boldsymbol{t}'\}\in P_k} \mathrm{Cov}[(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2, (Z(\boldsymbol{s}') - Z(\boldsymbol{t}'))^2] \\
&\leq \frac{1}{4}[(1-\epsilon)(32L\rho^*(C_k)) + 32L\epsilon] \\
&= 8L[(1-\epsilon)\rho^*(C_k) + \epsilon] \\
&\leq 8L(\rho^*(C_k) + \epsilon)
\end{aligned}$$

Applying Theorem 3.12, this gives

$$\begin{aligned}
\mathrm{Var}(\hat{\gamma}_{\delta_k}(h)) &\leq 32L\left(\rho^*(C_k) + D_3\frac{(h + \delta_k + C_k + 2\epsilon_k\sqrt{n})^n}{\lambda(V_k)}\right) \\
&\leq 32L\left(\rho^*(C_k) + D_3\frac{(h_0 + \delta_k + C_k + 2\epsilon_k\sqrt{n})^n}{\lambda(V_k)}\right) \\
&\leq D_4\left(\rho^*(C_k) + \left(\frac{C_k}{r_k}\right)^n\right)
\end{aligned}$$

for a suitable constant D_4. □

Theorem 3.4. *If Assumptions 3.2–3.6 and 3.8 are satisfied, then for all $h_0 > 0$, $\hat{\gamma}_{\delta_k}(h)$ converges uniformly to $\gamma(h)$ in probability on $[0, h_0]$, i.e., for any $\epsilon > 0$,*

$$\lim_{k\to\infty} P\left(\sup_{h\in[0,h_0]} |\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon\right) = 0$$

Proof. Since $\gamma(h)$ is uniformly continuous on bounded intervals, there exists $\delta > 0$ such that for all $h_1, h_2 \in \mathbb{R}$, if $|h_1 - h_2| < \delta$ and at least one of h_1 or h_2 is in $[0, h_0]$, then $|\gamma(h_1) - \gamma(h_2)| < \epsilon/3$. Choose finite subsets $T_k \subseteq [0, h_0]$ such that every point of $[0, h_0]$ has distance less than τ_k to a point of T_k, and choose T_k so that $|T_k| \le h_0/\tau_k$. (For example, the points of T_k could be spaced uniformly.)

Given any $h \in [0, h_0]$, if we let $h' \in T_k$ be the nearest point to h in T_k, then we have

$$|\hat{\gamma}_{\delta_k}(h) - \gamma(h)| \le |\hat{\gamma}_{\delta_k}(h) - \hat{\gamma}_{\delta_k}(h')| + |\hat{\gamma}_{\delta_k}(h') - \gamma(h')| + |\gamma(h') - \gamma(h)|$$

For k sufficiently large, we will have $\tau_k < \delta$, hence $|h' - h| < \delta$ and therefore, $|\gamma(h') - \gamma(h)| < \epsilon/3$. Thus for sufficiently large k, we have

$$\begin{aligned}
&P\left(\sup_{h\in[0,h_0]} |\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon\right) \\
&\le P\left(\sup_{|h-h'|<\tau_k} |\hat{\gamma}_{\delta_k}(h) - \hat{\gamma}_{\delta_k}(h')| > \epsilon/3\right) + P\left(\sup_{h\in T_k} |\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon/3\right)
\end{aligned}$$

To complete the proof, it suffices to show that each of these two terms converges to zero as $k \to \infty$. Since $E\hat{\gamma}_{\delta_k}(h) = \gamma(h)$, Chebyshev's inequality implies that for any $h \ge 0$,

$$P(|\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon/3) \le \frac{\mathrm{Var}(\hat{\gamma}_{\delta_k}(h))}{(\epsilon/3)^2}$$

Therefore, applying Boole's inequality and Theorem 3.13,

$$\begin{aligned}
P\left(\sup_{h\in T_k} |\hat{\gamma}_{\delta_k}(h) - \gamma(h)| > \epsilon/3\right) &\le |T_k| \frac{\mathrm{Var}(\hat{\gamma}_{\delta_k}(h))}{(\epsilon/3)^2} \\
&\le \frac{9h_0}{\tau_k \epsilon^2} D_4 \left(\rho^*(C_k) + \left(\frac{C_k}{r_k}\right)^n\right)
\end{aligned}$$

which converges to 0 by Assumption 3.8(a).

Now, for a given $[h_1, h_2]$ with $|h_1 - h_2| < \tau_k$, define the following subsets of pairs of points $(\boldsymbol{s}, \boldsymbol{t}) \in S_k \times S_k$:

$$\begin{aligned}
P_k = P_k(h_1, h_2) &= \{(\boldsymbol{s}, \boldsymbol{t}) \in S_k \times S_k \mid h_1 - \delta_k \le \|\boldsymbol{s} - \boldsymbol{t}\| < h_2 - \delta_k\} \\
Q_k = Q_k(h_1, h_2) &= \{(\boldsymbol{s}, \boldsymbol{t}) \in S_k \times S_k \mid h_2 - \delta_k \le \|\boldsymbol{s} - \boldsymbol{t}\| \le h_1 + \delta_k\} \\
R_k = R_k(h_1, h_2) &= \{(\boldsymbol{s}, \boldsymbol{t}) \in S_k \times S_k \mid h_1 + \delta_k < \|\boldsymbol{s} - \boldsymbol{t}\| \le h_2 + \delta_k\}
\end{aligned}$$

Since $\tau_k/\delta_k \to 0$ by Assumption 3.8(d), for sufficiently large k we know that $\tau_k < 2\delta_k$, so that the interval $[h_2 - \delta_k, h_1 + \delta_k]$ is nonempty. Given a pair $\boldsymbol{u} = (\boldsymbol{s}, \boldsymbol{t}) \in S_k \times S_k$, write $Z_{\boldsymbol{u}} = \frac{1}{2}(Z(\boldsymbol{s}) - Z(\boldsymbol{t}))^2$. Now we may write

$$\hat{\gamma}_{\delta_k}(h_1) = \frac{1}{|P_k \cup Q_k|} \sum_{\boldsymbol{u} \in P_k \cup Q_k} Z_{\boldsymbol{u}}$$
$$\hat{\gamma}_{\delta_k}(h_2) = \frac{1}{|Q_k \cup R_k|} \sum_{\boldsymbol{u} \in Q_k \cup R_k} Z_{\boldsymbol{u}}$$

Therefore, writing $p_k = |P_k|$, $q_k = |Q_k|$, and $r_k = |R_k|$, we have

$$\begin{aligned}
&|\hat{\gamma}_{\delta_k}(h_1) - \hat{\gamma}_{\delta_k}(h_2)| \\
&= \left| \frac{1}{p_k + q_k} \sum_{\boldsymbol{u} \in P_k} Z_{\boldsymbol{u}} + \left(\frac{1}{p_k + q_k} - \frac{1}{q_k + r_k} \right) \sum_{\boldsymbol{u} \in Q_k} Z_{\boldsymbol{u}} - \frac{1}{q_k + r_k} \sum_{\boldsymbol{u} \in R_k} Z_{\boldsymbol{u}} \right|
\end{aligned}$$

Now, with probability at least $1 - |S_k| P(Z(\boldsymbol{s}) > \beta_k)$, we have $|Z(\boldsymbol{s})| \leq \beta_k$ for all $\boldsymbol{s} \in S_k$, hence with at least the same probability we have $|Z_{\boldsymbol{u}}| \leq 2\beta_k^2$ for all $\boldsymbol{u} \in S_k \times S_k$, in which case

$$\begin{aligned}
&|\hat{\gamma}_{\delta_k}(h_1) - \hat{\gamma}_{\delta_k}(h_2)| \\
&\leq 2\beta_k^2 \left(\frac{p_k}{p_k + q_k} + \left| \frac{1}{p_k + q_k} - \frac{1}{q_k + r_k} \right| q_k + \frac{r_k}{q_k + r_k} \right) \\
&\leq 2\beta_k^2 \left(\frac{p_k}{p_k + q_k} + \frac{|r_k - p_k| q_k}{(p_k + q_k)(q_k + r_k)} + \frac{r_k}{q_k + r_k} \right)
\end{aligned}$$

To show that this quantity is less than $\epsilon/3$ for sufficiently large k, it suffices to show that $\beta_k^2 \frac{p_k}{q_k} \to 0$ and $\beta_k^2 \frac{r_k}{q_k} \to 0$ as $k \to \infty$, uniformly over all $[h_1, h_2] \subseteq [0, h_0]$ with $|h_1 - h_2| < \tau_k$, and this will complete the proof, since by Assumption 3.8(b) the probability $|S_k| P(|Z(\boldsymbol{s})| > \beta_k)$ converges to 0 as $k \to \infty$. Applying Lemma 3.11, for sufficiently large k, if $h_1 \geq \delta_k$ we have

$$\begin{aligned}
\beta_k^2 \frac{p_k}{q_k} &\leq \beta_k^2 \frac{N_k([h_1 - \delta_k, h_2 - \delta_k))}{N_k([h_2 - \delta_k, h_1 + \delta_k])} \\
&\leq \beta_k^2 \frac{D_2 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 - \delta_k + 2\epsilon_k \sqrt{n}\right)^{n-1} \left(h_2 - h_1 + 4\epsilon_k \sqrt{n}\right)}{D_1 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 - \delta_k + 2\epsilon_k \sqrt{n}\right)^{n-1} \left(2\delta_k - (h_2 - h_1) - 4\epsilon_k \sqrt{n}\right)} \\
&\leq \beta_k^2 \frac{D_2}{D_1} \frac{\tau_k + 4\epsilon_k \sqrt{n}}{2\delta_k - \tau_k - 4\epsilon_k \sqrt{n}} \\
&= \frac{D_2}{D_1} \frac{\beta_k^2 \frac{\tau_k}{\delta_k} + 4\beta_k^2 \frac{\epsilon_k}{\delta_k} \sqrt{n}}{2 - \frac{\tau_k}{\delta_k} - 4\sqrt{n} \frac{\epsilon_k}{\delta_k}}
\end{aligned}$$

and the last expression does not depend on h_1 and h_2 and converges to 0 by Assumption 3.8(c,d). If $h_2 < \delta_k$, then $p_k = 0$, in which case $\beta_k^2 \frac{p_k}{q_k} = 0$. In the remaining case, where $h_1 < \delta_k \leq h_2$, for sufficiently large k we have

$$\begin{aligned}
\beta_k^2 \frac{p_k}{q_k} &\leq \beta_k^2 \frac{N_k([h_1 - \delta_k, h_2 - \delta_k))}{N_k([h_2 - \delta_k, h_1 + \delta_k])} \\
&= \beta_k^2 \frac{N_k([0, h_2 - \delta_k))}{N_k([h_2 - \delta_k, h_1 + \delta_k])} \\
&\leq \beta_k^2 \frac{D_2 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h_2 - \delta_k + 2\epsilon_k\sqrt{n})\right)}{D_1 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 - \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(2\delta_k - (h_2 - h_1) - 4\epsilon_k\sqrt{n}\right)} \\
&= \beta_k^2 \frac{D_2 \frac{|S_k|^2}{\lambda(V_k)} u (h_2 - \delta_k + 2\epsilon_k\sqrt{n})^n}{D_1 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 - \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(2\delta_k - (h_2 - h_1) - 4\epsilon_k\sqrt{n}\right)} \\
&= \beta_k^2 \frac{D_2}{nD_1} \frac{h_2 - \delta_k + 2\epsilon_k\sqrt{n}}{2\delta_k - (h_2 - h_1) - 4\epsilon_k\sqrt{n}} \\
&\leq \beta_k^2 \frac{D_2}{nD_1} \frac{\tau_k + 2\epsilon_k\sqrt{n}}{2\delta_k - \tau_k - 4\epsilon_k\sqrt{n}} \\
&= \frac{D_2}{nD_1} \frac{\beta_k^2 \frac{\tau_k}{\delta_k} + 2\sqrt{n}\beta_k^2 \frac{\epsilon_k}{\delta_k}}{2 - \frac{\tau_k}{\delta_k} - 4\sqrt{n}\frac{\epsilon_k}{\delta_k}}
\end{aligned}$$

which again converges to 0 by Assumption 3.8(c,d). Similarly, for sufficiently large k, if $h_2 \geq \delta_k$,

$$\begin{aligned}
\beta_k^2 \frac{r_k}{q_k} &\leq \beta_k^2 \frac{N_k((h_1 + \delta_k, h_2 + \delta_k])}{N_k([h_2 - \delta_k, h_1 + \delta_k])} \\
&\leq \beta_k^2 \frac{D_2 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(h_2 - h_1 + 4\epsilon_k\sqrt{n}\right)}{D_1 \frac{|S_k|^2}{\lambda(V_k)} u \left(h_1 + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(2\delta_k - (h_2 - h_1) - 4\epsilon_k\sqrt{n}\right)} \\
&= \beta_k^2 \frac{D_2 n}{D_1} \left(1 + \frac{h_2 - h_1}{h_1 + \delta_k + 2\epsilon_k\sqrt{n}}\right)^{n-1} \frac{h_2 - h_1 + 4\epsilon_k\sqrt{n}}{2\delta_k - (h_2 - h_1) - 4\epsilon_k\sqrt{n}} \\
&\leq \beta_k^2 \frac{D_2 n}{D_1} \left(1 + \frac{\tau_k}{h_2 - \tau_k + \delta_k + 2\epsilon_k\sqrt{n}}\right)^{n-1} \frac{\tau_k + 4\epsilon_k\sqrt{n}}{2\delta_k - \tau_k - 4\epsilon_k\sqrt{n}} \\
&\leq \beta_k^2 \frac{D_2 n}{D_1} \left(1 + \frac{\tau_k}{\delta_k - \tau_k + \delta_k + 2\epsilon_k\sqrt{n}}\right)^{n-1} \frac{\tau_k + 4\epsilon_k\sqrt{n}}{2\delta_k - \tau_k - 4\epsilon_k\sqrt{n}} \\
&= \frac{D_2 n}{D_1} \left(1 + \frac{\frac{\tau_k}{\delta_k}}{2 - \frac{\tau_k}{\delta_k} + 2\sqrt{n}\frac{\epsilon_k}{\delta_k}}\right)^{n-1} \frac{\beta_k^2 \frac{\tau_k}{\delta_k} + 4\sqrt{n}\beta_k^2 \frac{\epsilon_k}{\delta_k}}{2 - \frac{\tau_k}{\delta_k} - 4\sqrt{n}\frac{\epsilon_k}{\delta_k}}
\end{aligned}$$

which converges to 0 by Assumption 3.8(c,d). For $h_2 < \delta_k$, we have

$$\begin{aligned}
\beta_k^2 \frac{r_k}{q_k} &\leq \beta_k^2 \frac{N_k([h_1 + \delta_k, h_2 + \delta_k])}{N_k([h_2 - \delta_k, h_1 + \delta_k])} \\
&\leq \beta_k^2 \frac{N_k([h_1 + \delta_k, h_2 + \delta_k])}{N_k([0, h_1 + \delta_k])}
\end{aligned}$$

$$
\begin{aligned}
&\leq \beta_k^2 \frac{D_2 \frac{|S_k|^2}{\lambda(V_k)} nu \left(h_2 + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(h_2 - h_1 + 4\epsilon_k\sqrt{n}\right)}{D_1 \frac{|S_k|^2}{\lambda(V_k)} \lambda\left(B_{\mathbf{0}}(h_1 + \delta_k - 2\epsilon_k\sqrt{n})\right)} \\
&= \beta_k^2 \frac{D_2 nu \left(h_2 + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(h_2 - h_1 + 4\epsilon_k\sqrt{n}\right)}{D_1 u (h_1 + \delta_k - 2\epsilon_k\sqrt{n})^n} \\
&\leq \beta_k^2 \frac{D_2 n \left(\delta_k + \delta_k + 2\epsilon_k\sqrt{n}\right)^{n-1} \left(\tau_k + 4\epsilon_k\sqrt{n}\right)}{D_1 (\delta_k - 2\epsilon_k\sqrt{n})^n} \\
&= \frac{D_2 n \left(2 + 2\frac{\epsilon_k}{\delta_k}\sqrt{n}\right)^{n-1} \left(\beta_k^2 \frac{\tau_k}{\delta_k} + 4\beta_k^2 \frac{\epsilon_k}{\delta_k}\sqrt{n}\right)}{D_1 (1 - 2\frac{\epsilon_k}{\delta_k}\sqrt{n})^n}
\end{aligned}
$$

which again converges to 0 by Assumption 3.8(c,d). □

Theorem 3.5. *Suppose that* $E|Z(\boldsymbol{s}) - \mu|^p < \infty$, *and that* $\rho^*(h) \ll h^{-\alpha_\rho}$, $r_k \asymp k^{\alpha_r}$, $|S_k| \asymp k^{\alpha_m}$, $\delta_k \asymp k^{-\alpha_\delta}$, $\epsilon_k \asymp k^{-\alpha_\epsilon}$, $C_k \asymp k^{\alpha_C}$, $\tau_k \asymp k^{-\alpha_\tau}$, *and* $\beta_k \asymp k^{\alpha_\beta}$ *for some constants* $p, \alpha_\rho, \alpha_r, \alpha_m, \alpha_\delta, \alpha_\epsilon, \alpha_C, \alpha_\tau, \alpha_\beta > 0$. *Then Assumption 3.8 holds provided that*

(a) $\alpha_\tau < \min\{\alpha_C\alpha_\rho, n(\alpha_r - \alpha_C)\}$,

(b) $\alpha_m < \alpha_\beta p$,

(c) $2\alpha_\beta < \alpha_\epsilon - \alpha_\delta$,

(d) $2\alpha_\beta < \alpha_\tau - \alpha_\delta$.

Proof. (a) We have

$$
\frac{\rho^*(C_k)}{\tau_k} \ll \frac{C_k^{-\alpha_\rho}}{k^{\alpha_\tau}} \asymp \frac{k^{-\alpha_C\alpha_\rho}}{k^{-\alpha_\tau}} = k^{\alpha_\tau - \alpha_C\alpha_\rho}
$$

which converges to 0 provided $\alpha_\tau < \alpha_C\alpha_\rho$. Also,

$$
\frac{\left(\frac{C_k}{r_k}\right)^n}{\tau_k} \asymp k^{n(\alpha_C - \alpha_r) + \alpha_\tau}
$$

which converges to 0 provided $\alpha_\tau < n(\alpha_r - \alpha_C)$.

(b) Let $M = E|Z(\boldsymbol{s}) - \mu|^p$. Then by Markov's inequality,

$$
\begin{aligned}
|S_k| P(|Z(\boldsymbol{s}) - \mu| > \beta_k) &= |S_k| P(|Z(\boldsymbol{s}) - \mu|^p > \beta_k^p) \\
&\leq |S_k| M / \beta_k^p \\
&\asymp k^{\alpha_m - \alpha_\beta p}
\end{aligned}
$$

which converges to 0 provided $\alpha_m < \alpha_\beta p$.

(c) We have

$$\beta_k^2 \epsilon_k / \delta_k \asymp k^{2\alpha_\beta - \alpha_\epsilon + \alpha_\delta}$$

which converges to 0 provided $2\alpha_\beta < \alpha_\epsilon - \alpha_\delta$.

(d) We have

$$\beta_k^2 \tau_k / \delta_k \asymp k^{2\alpha_\beta - \alpha_\tau + \alpha_\delta}$$

which converges to 0 provided $2\alpha_\beta < \alpha_\tau - \alpha_\delta$.

□

Theorem 3.6. *Suppose that $E|Z(\boldsymbol{s}) - \mu|^p < \infty$ for some $p > 2n$, and that $\rho^*(h) \ll h^{-\alpha_\rho}$, $r_k \asymp k^{\alpha_r}$, $m_k \asymp k^{\alpha_m}$, $\delta_k \asymp k^{-\alpha_\delta}$, for some constants $\alpha_\rho, \alpha_r, \alpha_m, \alpha_\delta > 0$, and suppose further that*

(a) $\dfrac{\alpha_m}{n\alpha_r} < \dfrac{p}{2}\left(\dfrac{\alpha_\rho}{\alpha_\rho + n} - \dfrac{\alpha_\delta}{n\alpha_r}\right)$

(b) $\alpha_\delta < \alpha_m\left(\dfrac{1}{n} - \dfrac{2}{p}\right) - \alpha_r$

Then constants $\alpha_\epsilon, \alpha_C, \alpha_\tau, \alpha_\beta$ exist which satisfy Assumption 3.7 as well as the assumptions of Theorem 3.5, if locations sets S_k are chosen with $|S_k| = m_k$.

Proof. Choose α_β to be any number such that

$$\frac{\alpha_m}{p} < \alpha_\beta < \min\left\{\frac{n\alpha_r\alpha_\rho}{2(\alpha_\rho + n)}, \frac{\alpha_m}{2n} - \frac{\alpha_r}{2}\right\} - \frac{\alpha_\delta}{2}, \tag{3.5}$$

which is possible by assumption (a) and by assumption (b). The first inequality in (3.5) implies that assumption (b) of Theorem 3.5 is satisfied. Now set $\alpha_C = \frac{n\alpha_r}{\alpha_\rho + n}$, so that $\alpha_C\alpha_\rho = n(\alpha_r - \alpha_C)$. Now choose α_τ to be any number such that $2\alpha_\beta + \alpha_\delta < \alpha_\tau < \alpha_C\alpha_\rho$, which is possible since the second inequality in (3.5) implies $2\alpha_\beta < \alpha_C\alpha_\rho - \alpha_\delta$. This ensures that assumptions (a) and (d) of Theorem 3.5 are satisfied. Finally, choose α_ϵ to be any number such that $2\alpha_\beta + \alpha_\delta < \alpha_\epsilon < \alpha_m/n - \alpha_r$, which is possible since we chose α_β such that $\alpha_\beta < \alpha_m/2n - \alpha_r/2 - \alpha_\delta/2$. The inequality $2\alpha_\beta + \alpha_\delta < \alpha_\epsilon$ then implies assumption (c) of Theorem 3.5. Finally, the inequality $\alpha_\epsilon < \alpha_m/n - \alpha_r$ implies Assumption 3.7. □

Theorem 3.7. *Suppose that Assumptions (3.1) and (3.9)–(3.12) are satisfied. If C_k and δ_k are sequences of positive numbers such that $\rho^*(C_k) \to 0$, $\delta_k \to 0$, and $p_k(h + C_k + \delta_k; h, \delta_k) \to 0$, then $\hat{\gamma}_{ij,\delta_k}(h)$ is consistent.*

Proof. The proof follows the same structure as that of Theorem 3.1. Let $L = \sup_i E(Z_i(s)^4)$, so $L < \infty$ by Assumption 3.11. We note that for any $i = 1, \ldots, p$ and any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^n$, we

have $E[(Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t}))^4] \leq 32L$. Fix $\epsilon > 0$ and $\rho > 0$, and assume that k is sufficiently large so that $p_k(h + C_k + \delta_k; h, \delta_k) < \epsilon$ and $\rho^*(C_k) < \rho$, by Assumptions 3.1 and 3.12. Given a pair of points $\boldsymbol{s}, \boldsymbol{t} \in S_k$ with $h - \delta_k < |\boldsymbol{s} - \boldsymbol{t}| < h + \delta_k$, taking $U = B_{\boldsymbol{s}}(h + \delta_k + C_k)$ we know that at least proportion $1 - \epsilon$ of the pairs $\boldsymbol{s}', \boldsymbol{t}'$ with $h - \delta < |\boldsymbol{s}' - \boldsymbol{t}'| < h + \delta$ have $d(\{\boldsymbol{s}, \boldsymbol{t}\}, \{\boldsymbol{s}', \boldsymbol{t}'\}) \geq C_k$, and for these, we may bound the covariance

$$
\begin{aligned}
&\operatorname{Cov}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})], [Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')]) \\
&= \operatorname{Corr}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})], [Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')]) \\
&\quad \cdot \sqrt{\operatorname{Var}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})]) \operatorname{Var}([Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')])} \\
&\leq \rho \sqrt{\operatorname{Var}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})]) \operatorname{Var}([Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')])} \\
&\leq \rho \sqrt{E([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})]^2 [Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})]^2) E([Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')]^2 [Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')]^2)} \\
&\leq \rho \sqrt{\sqrt{E[(Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t}))^4]} \sqrt{E[(Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t}))^4]}} \\
&\quad \cdot \sqrt{\sqrt{E[(Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}'))^4]} \sqrt{E[(Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}'))^4]}} \\
&\leq 32L\rho
\end{aligned}
$$

For the remaining pairs $\boldsymbol{s}', \boldsymbol{t}'$ we have the looser bound

$$
\begin{aligned}
&\operatorname{Cov}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})], [Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')]) \\
&\leq \sqrt{\operatorname{Var}([Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})]) \operatorname{Var}([Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')])} \\
&\leq 32L
\end{aligned}
$$

As above, define P_k to be the set of pairs of points $\{\boldsymbol{s}, \boldsymbol{t}\} \in S_k$ with lag in the prescribed range, i.e., $h - \delta_k < |\boldsymbol{s} - \boldsymbol{t}| < h + \delta_k$. We then have a bound

$$
\begin{aligned}
\operatorname{Var}(\hat{\gamma}_{ij,\delta}) &= \frac{1}{4N(h)^2} \operatorname{Var}\left[\sum_{\{\boldsymbol{s},\boldsymbol{t}\} \in P_k} [Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})] \right] \\
&= \frac{1}{4N(h)^2} \sum_{\{\boldsymbol{s},\boldsymbol{t}\} \in P_k} \sum_{\{\boldsymbol{s}',\boldsymbol{t}'\} \in P_k} \operatorname{Cov}\begin{pmatrix} [Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})], \\ [Z_i(\boldsymbol{s}') - Z_i(\boldsymbol{t}')][Z_j(\boldsymbol{s}') - Z_j(\boldsymbol{t}')] \end{pmatrix} \\
&\leq \frac{1}{4}[(1 - \epsilon)(32\rho L) + 32L\epsilon] \\
&= 8L[(1 - \epsilon)\rho + \epsilon]
\end{aligned}
$$

For fixed δ, letting $\rho \to 0$ and $\epsilon \to 0$ we thus have $\operatorname{Var}(\hat{\gamma}_{ij,\delta}) \to 0$. It only remains to show that the estimator is asymptotically unbiased. The continuity of γ_{ij} (Assumption

3.10) ensures that for any $\epsilon > 0$ there is some $\delta_0 > 0$ such that $|\gamma_{ij}(h+\delta) - \gamma_{ij}(h)| \leq \epsilon$ for all $\delta \leq \delta_0$. Hence for k sufficiently large so that $\delta_k < \delta_0$,

$$\begin{aligned} E[\hat{\gamma}_{ij,\delta_k}] &= \frac{1}{N(H)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\} \in P_k} E[Z_i(\boldsymbol{s}) - Z_i(\boldsymbol{t})][Z_j(\boldsymbol{s}) - Z_j(\boldsymbol{t})] \\ &= \frac{1}{N(H)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\} \in P_k} \gamma_{ij}(|\boldsymbol{s} - \boldsymbol{t}|) \\ &\leq \frac{1}{N(H)} \sum_{\{\boldsymbol{s},\boldsymbol{t}\} \in P_k} (\gamma_{ij}(h) + \epsilon) = \gamma_{ij}(h) + \epsilon \end{aligned}$$

and similarly $E[\hat{\gamma}_{ij,\delta_k}] \geq \gamma_{ij}(h) - \epsilon$. But as $\delta_i \to 0$ we may take $\epsilon \to 0$, completing the proof. □

CHAPTER 4

APPLICATIONS

4.1 Image compression

Many compressed image formats, including the popular JPEG format, are based on applying the two-dimensional discrete cosine transform (DCT) to small, square blocks of the image. Other formats, including JPEG 2000, are based on a discrete wavelet transform (DWT). In both cases, the idea is that the transform tends to concentrate most of the useful information into a small number of coefficients, corresponding to the low-frequency (smooth) components of the image. Then, in a step known as *quantization*, the coefficients are rounded to a discrete set of possible values, with spacing determined by the desired trade-off between quality and file size. One goal is that in this step many of the high-frequency coefficients will round to zero. Finally, the discretized coefficients are processed by *entropy coding*, which uses a probability model to encode more likely values, such as zero, using few bits (possibly even less than a whole bit), while encoding less likely values using more bits.

It has been recognized, however, that the DCT and DWT are suboptimal in capturing anisotropic features commonly found in images, and this has spurred efforts to develop better-performing methods [1, 40, 29, 12]. In this section, we will consider a new approach to this problem based on anisotropic spatial statistical modeling. For simplicity, we consider grayscale images, where each pixel is represented by a real number between 0 (black) and 1 (white). As with other methods, this approach has three parts: transformation, quantization, and entropy coding. The main difference is that instead of using a fixed transform, the choice of transform is determined dynamically by the estimated parameters of a spatial model.

Since many images feature distinct objects each with varying appearances, it would be inappropriate to assume stationarity across the entirety of an image. Instead, the image is partitioned into blocks of 16×16 pixels and stationarity is assumed on each block. Each block is then modeled as a realization of a random field $Z(A\boldsymbol{s})$ for some 2×2 upper triangular matrix A, where $Z(\boldsymbol{s})$ is a stationary, isotropic random field with semivariogram $\gamma(h) = \sigma^2(1 - e^{-h})$, and the data are modeled as observations of the random field $Z(A\boldsymbol{s})$

at locations $\boldsymbol{s} = (i, j)$ for $i, j \in \{1, \ldots, 16\}$. The matrix A introduces anisotropy into the model; if A is a scalar matrix λI, then the resulting model becomes isotropic. Within each block, quantized estimates $\hat{A}$ and $\hat{\sigma}^2$ of the parameters A and σ^2 are computed; we do this by enumerating over all possible values for $\hat{A}$ and $\hat{\sigma}^2$ (within a small finite set of quantized values which are fixed in advance) and choosing the combination which results in the smallest entropy-coded representation of the block (in the step described in the next paragraph). This method of estimating A and σ^2 is computationally intensive but is sufficient for the purposes of illustrating the algorithm in its most basic form.

Now, we let X be the random vector in $\mathbb{R}^{256}$ representing the data in the given 16×16 block. An estimate $\hat{\Sigma}$ of the covariance matrix $\Sigma = \mathrm{Var}(X)$ is determined based on the estimates $\hat{A}$ and $\hat{\sigma}^2$ and the assumed spatial model. By the spectral theorem, we may write $\hat{\Sigma} = QDQ^T$, where Q is an orthogonal matrix and D is a diagonal matrix with nonnegative entries. Now let $Y = Q^T X$, which is the Karhunen-Loeve transform (KLT) of X based on the estimated covariance matrix. If $\hat{\Sigma}$ were equal to the true covariance matrix, then the components of Y would be uncorrelated. We model the components of Y as independent Laplace random variables and perform optimal entropy-constrained quantization using the method of Sullivan [39]. Finally, the discretized components are processed by the binary entropy coding method of Duda [13], yielding the compressed output. The quantized estimated model parameters $\hat{A}$ and $\hat{\sigma}^2$ for each block are also included in the output.

The application of this KLT method on a sample image is shown in Figure 4.1; its favorable performance compared to DCT- and DWT-based methods is shown in Figure 4.2, where the quality of the encoded images is measured by the peak signal-to-noise ratio (PSNR), defined in decibel (dB) units as

$$\mathrm{PSNR} = 10 \log_{10}(1/\mathrm{MSE})$$

where MSE is the mean squared error between the encoded image and the original image. The partitioning and transformation of the sample image is shown in Figure 4.3, and two of the KLT bases are shown in Figure 4.4.

A number of refinements could potentially improve the method further: instead of assuming that the blocks are independent of one another, their dependence could be exploited in the form of interblock prediction; variable-sized blocks could be used instead of a fixed size of 16×16; additional types of spatial models could be considered; and an optimized method could be developed for the quantization of the estimated model parameters.

Figure 4.1: Original and compressed versions of `Barbara` image. Left: Original image (262 kB). Right: Image compressed by KLT method (27 kB).

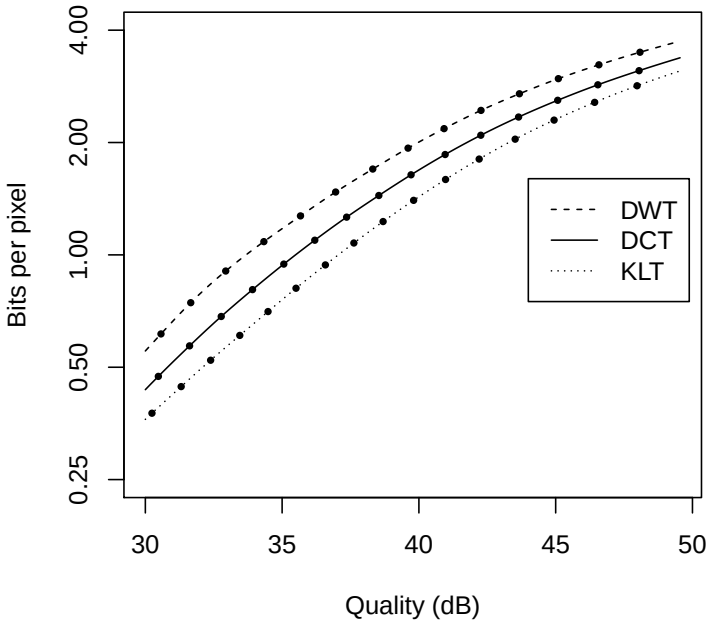

Figure 4.2: Comparison of DWT, DCT, and KLT methods on the `Barbara` image.

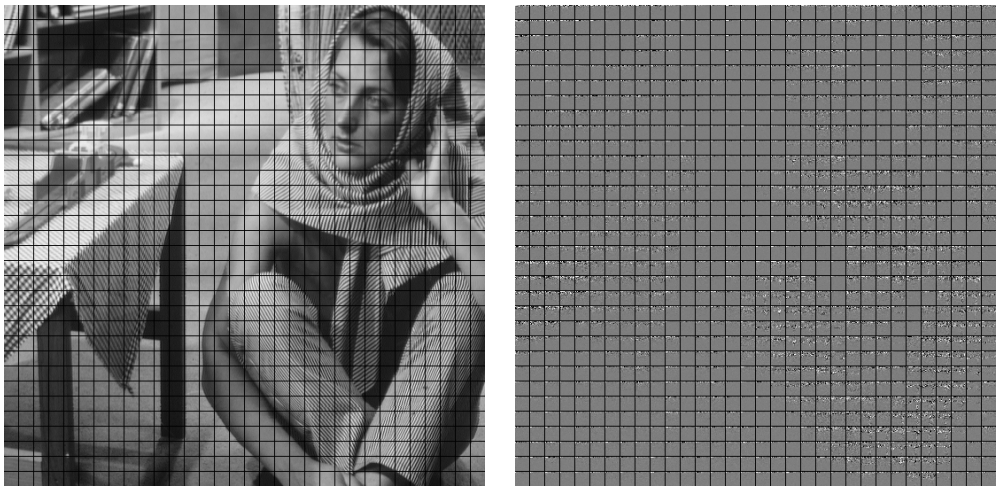

Figure 4.3: Partitioned version of original and transformed image. Left: Image partitioned in 16×16 blocks. Right: KLT transforms of blocks.

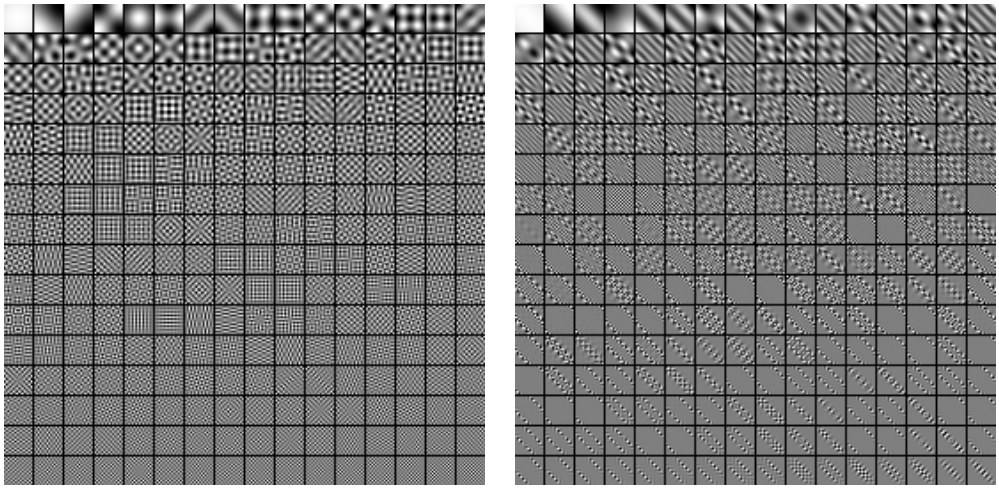

Figure 4.4: Karhunen-Loéve bases for two transformations. Left: Karhunen-Loéve basis for $A = \begin{pmatrix} 0.0312 & 0 \\ 0 & 0.0312 \end{pmatrix}$, giving an isotropic model. Right: Karhunen-Loéve basis for $A = \begin{pmatrix} -0.0231 & 0.0191 \\ 0 & 0.0130 \end{pmatrix}$, giving an anisotropic model.

4.2 Moran's I test for autocorrelation

In this section, we consider Moran's I test for spatial autocorrelation. In this setting, we observe random variables $Z_1, \ldots, Z_m$, representing data collected from m different spatial locations or regions, and we test the null hypothesis that $Z_1, \ldots, Z_m$ are independent and identically distributed. The test rejects the null hypothesis for large values of Moran's I statistic:

$$I = \frac{m \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij} (Z_i - \overline{Z})(Z_j - \overline{Z})}{(m-1) S^2 \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij}} \tag{4.1}$$

where w_{ij} are "weights" which must be chosen by the practitioner. Without loss of generality, it may be assumed that the matrix of weights is symmetric, i.e., $w_{ij} = w_{ji}$. If the data represent aggregates over spatial regions (such as counties), a popular choice of weights is to set $w_{ij} = 1$ if the regions i and j are contiguous and to set $w_{ij} = 0$ otherwise. In this case, the test statistic I can be understood as measuring the extent to which data in adjacent counties are correlated. On the other hand, in the case where the data $Z_1, \ldots, Z_m$ represent observations of a spatial random field at given locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$, one popular choice is to set $w_{ij} = 1/d_{ij}$, where d_{ij} is the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. Such choices of weights appear to be ad-hoc, motivating us to consider an alternative approach.

We first note that Moran's I statistic may be expressed more concisely as

$$I = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} u_{ij} Z_i Z_j}{S^2} \tag{4.2}$$

for a suitable choice of u_{ij} satisfying $u_{ij} = u_{ji}$ and $\sum_{j=1}^{m} u_{ij} = 0$ for each i, namely

$$u_{ij} = \frac{m}{(m-1) \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij}} \left(w_{ij} - \frac{1}{m} \sum_{k=1}^{m} w_{ik} - \frac{1}{m} \sum_{k=1}^{m} w_{kj} + \frac{1}{m^2} \sum_{k=1}^{m} \sum_{l=1}^{m} w_{kl} \right) \tag{4.3}$$

Conversely, given any real numbers u_{ij} satisfying $u_{ij} = u_{ji}$ and $\sum_{j=1}^{m} u_{ij} = 0$ for each i, if we rescale u_{ij} by a suitable constant factor, then (4.1) holds with $w_{ij} = u_{ij}$. We observe that the distribution of I is unchanged by shifting or rescaling $Z_1, \ldots, Z_m$ by a common constant. Therefore, if we assume that Z_i has a finite second moment, then there is no further loss in generality in assuming that $Z_1, \ldots, Z_m$ each have mean 0 and variance 1. Let Y be the numerator in (4.2):

$$Y = \sum_{i=1}^{m} \sum_{j=1}^{m} u_{ij} Z_i Z_j$$

In selecting the weights u_{ij} with the goal of achieving a powerful test, one strategy is to maximize the expected value of the standardization of Y; in other words, we consider the following criterion:

Criterion. *Choose u_{ij} so as to maximize the expected value of $\frac{Y-\mu_Y}{\sigma_Y}$ under the alternative hypothesis, where μ_Y and σ_Y are the mean and standard deviation of Y under the null hypothesis that $Z_1, \ldots, Z_m$ are independent random variables each with mean 0 and variance 1.*

If we assume that $Z_1, \ldots, Z_m$ follow a multivariate Gaussian distribution, and if the hypotheses are fully specified, then we will show that this criterion becomes well-defined, in the following sense:

Theorem 4.1. *Given an alternative hypothesis that $Z_1, \ldots, Z_m$ are identically distributed (but not necessarily independent), having a multivariate Gaussian distribution with a specified covariance matrix $\Sigma = (\sigma_{ij})$, the criterion above determines a unique choice of u_{ij} up to multiplication by a constant, namely*

$$u_{ij} = \begin{cases} -\dfrac{2}{m}\displaystyle\sum_{k\neq i}\sigma_{ik} + \dfrac{1}{m^2}\sum_{k=1}^{m}\sum_{l\neq k}\sigma_{kl}, & i = j, \\ \sigma_{ij} - \dfrac{1}{m}\displaystyle\sum_{k\neq i}\sigma_{ik} - \dfrac{1}{m}\sum_{k\neq j}\sigma_{jk} + \dfrac{1}{m^2}\sum_{k=1}^{m}\sum_{l\neq k}\sigma_{kl}, & i \neq j. \end{cases}$$

In terms of (4.1), up to a constant factor the corresponding weights are simply $w_{ij} = \sigma_{ij}$ for $i \neq j$, and $w_{ii} = 0$.

Another form of (4.2) is known as the Geary c statistic:

$$c = \frac{m\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(Z_i - Z_j)^2}{2S^2 \sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}} \tag{4.4}$$

Whereas (4.1) may be interpreted as a weighted sum of covariance estimators, (4.4) may be interpreted as a weighted sum of semivariogram estimators. In this case, the test rejects the null hypothesis for small values of c. Up to a negative constant factor, Moran's I statistic in (4.2) is identical to Geary's c statistic with $v_{ij} = u_{ij}$.

Given real spatial data, the covariance matrix $\Sigma = (\sigma_{ij})$ is generally unknown, but it may be estimated based on a parametric model, and the estimates $\hat{\sigma}_{ij}$ may be used for the weights w_{ij}. To measure the performance of this method compared with the two methods mentioned previously, we carry out simulations of Gaussian random fields, observed on a

square 10×10 grid of points $\{1, \ldots, 10\}^2$. The simulations are based on an exponential covariance function $C(h) = e^{-h/h_0}$ with three choices of range parameter, $h_0 = 0.4$, 0.7, and 1.0. Data are also simulated for the case of the null hypothesis $C(h) = 1_{\{0\}}(h)$, which for convenience we designate by $h_0 = 0$. The covariance function is estimated using the method of maximum likelihood, by the `likfit` in the R package `geoR`, based on the covariance model $C(h) = \sigma^2 e^{-h/h_0}$. We computed the Moran's I statistic and P-value using the function `lm.morantest.exact` in the `spdep` package. The results are shown in Table 4.1. In the case where the null hypothesis is true, the distance method (where $w_{ij} = 1/d_{ij}$) and the adjacency method (where $w_{ij} = 1$ if $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are adjacent and $w_{ij} = 0$ otherwise) both reject the null hypothesis at very close to the nominal rate, while the new "estimated covariance" method is somewhat conservative. In cases where the alternative hypothesis is true, the new method has higher power than the inverse distance method but lower power than the adjacency method.

Table 4.1: Estimated rejection rates of Moran's I test for three methods of choosing the weights w_{ij}, based on 1,355,715 simulations of Gaussian random fields at each of four range parameters.

Range	Nominal size	Rejection rate		
		Adjacency	Inverse distance	Estimated covariance
$h_0 = 0$	$\alpha = 0.1$	0.0996	0.0998	0.0964
	$\alpha = 0.05$	0.0497	0.0495	0.0474
	$\alpha = 0.01$	0.0100	0.0099	0.0091
	$\alpha = 0.005$	0.0050	0.0050	0.0046
	$\alpha = 0.001$	0.0010	0.0010	0.0009
$h_0 = 0.4$	$\alpha = 0.1$	0.4081	0.3444	0.4067
	$\alpha = 0.05$	0.2801	0.2234	0.2741
	$\alpha = 0.01$	0.1077	0.0760	0.0998
	$\alpha = 0.005$	0.0694	0.0468	0.0626
	$\alpha = 0.001$	0.0240	0.0148	0.0204
$h_0 = 0.7$	$\alpha = 0.1$	0.9278	0.8545	0.9263
	$\alpha = 0.05$	0.8736	0.7602	0.8635
	$\alpha = 0.01$	0.7103	0.5223	0.6661
	$\alpha = 0.005$	0.6326	0.4289	0.5726
	$\alpha = 0.001$	0.4576	0.2562	0.3732
$h_0 = 1$	$\alpha = 0.1$	0.9941	0.9772	0.9934
	$\alpha = 0.05$	0.9871	0.9513	0.9836
	$\alpha = 0.01$	0.9543	0.8471	0.9287
	$\alpha = 0.005$	0.9320	0.7870	0.8883
	$\alpha = 0.001$	0.8608	0.6321	0.7597

4.3 Proofs

Proof of Theorem 4.1. We calculate

$$\mu_Y = E_0(Y) = \sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}E(Z_iZ_j) = \sum_{i=1}^{m} u_{ii}E(Z_i^2) = \sum_{i=1}^{m} u_{ii}$$

$$\begin{aligned}
E_0(Y^2) &= E\left(\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}Z_iZ_j\right)^2 \\
&= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m}\sum_{l=1}^{m} u_{ij}u_{kl}E(Z_iZ_jZ_kZ_l) \\
&= \sum_{i=1}^{m} u_{ii}^2E(Z_i^4) + \sum_{i=1}^{m}\sum_{k\neq i} u_{ii}u_{kk}E(Z_i^2Z_k^2) + 2\sum_{i=1}^{m}\sum_{j\neq i}^{m} u_{ij}^2E(Z_i^2Z_j^2) \\
&= \sum_{i=1}^{m} 3u_{ii}^2 + \sum_{i=1}^{m}\sum_{k\neq i} u_{ii}u_{kk} + 2\sum_{i=1}^{m}\sum_{j\neq i}^{m} u_{ij}^2 \\
&= \sum_{i=1}^{m}\sum_{k=1}^{m} u_{ii}u_{kk} + 2\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}^2 = \mu_Y^2 + 2\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}^2
\end{aligned}$$

Therefore, under the null hypothesis,

$$\sigma_Y^2 = 2\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}^2$$

By rescaling all the u_{ij} by a constant factor, we may assume $\sum_{i,j} u_{ij}^2 = 1$, in which case

$$\sigma_Y^2 = 2$$

Under the alternative hypothesis, we calculate

$$E(Y) = \sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}E(Z_iZ_j) = \sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}\sigma_{ij}$$

Thus, under the alternative hypothesis

$$\begin{aligned}
E\left(\frac{Y-\mu_Y}{\sigma_Y}\right) &= \frac{1}{\sqrt{2}}\left(\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}\sigma_{ij} - \sum_{i=1}^{m} u_{ii}\right) \\
&= \frac{1}{\sqrt{2}}\sum_{i=1}^{m}\sum_{j\neq i} u_{ij}\sigma_{ij}
\end{aligned}$$

We wish to maximize this quantity, subject to the constraints, $u_{ij} = u_{ji}$, $\sum_{i=1}^{m} u_{ij} = 0$, and $\sum_{i=1}^{m} u_{ij}^2 = 1$. To simplify the computation, we consider only u_{ij} for $i < j$ as the

independent variables, and we may regard $u_{ij} = u_{ji}$ and $\sum_{i=1}^{m} u_{ij} = 0$ as equations defining u_{ij} for $j < i$ and $j = i$, respectively, leaving us with a single constraint $\sum_{i=1}^{m}\sum_{j=1}^{m} u_{ij}^2 = 1$, which may be rewritten

$$\sum_{i=1}^{m}\sum_{j\neq i} u_{ij}^2 + \sum_{i=1}^{m}\left(\sum_{j\neq i} u_{ij}\right)^2 = 1$$

The method of Lagrange multipliers gives us a system of equations

$$\frac{2}{\sqrt{2}}\sigma_{ij} = \lambda\left(4u_{ij} + 2\sum_{k\neq i}^{m} u_{ik} + 2\sum_{k\neq j}^{m} u_{jk}\right)$$

for $i \neq j$. Equivalently, for $i \neq j$,

$$\sigma_{ij} = \lambda(2u_{ij} - u_{ii} - u_{jj}) \tag{4.5}$$

Let $\alpha = \sum_{i=1}^{m} u_{ii}$. For fixed i, summing equation (4.5) over all $j \neq i$ gives

$$\sum_{j\neq i}\sigma_{ij} = \lambda(-2u_{ii} - (m-1)u_{ii} - \alpha + u_{ii}) = -\lambda(\alpha + mu_{ii}) \tag{4.6}$$

Summing this over all i gives

$$\sum_{i=1}^{m}\sum_{j\neq i}\sigma_{ij} = -\lambda(m\alpha + m\alpha) = -2\lambda m\alpha$$

Therefore,

$$\lambda\alpha = -\frac{1}{2m}\sum_{i=1}^{m}\sum_{j\neq i}\sigma_{ij}$$

Substituting into equation (4.6) gives

$$\begin{aligned}
\lambda u_{ii} &= -\frac{1}{m}\left(\sum_{j\neq i}\sigma_{ij} + \lambda\alpha\right) \\
&= -\frac{1}{m}\left(\sum_{j\neq i}\sigma_{ij} - \frac{1}{2m}\sum_{k=1}^{m}\sum_{l\neq k}\sigma_{kl}\right) \\
&= -\frac{1}{m}\sum_{j\neq i}\sigma_{ij} + \frac{1}{2m^2}\sum_{k=1}^{m}\sum_{l\neq k}\sigma_{kl}
\end{aligned}$$

Finally, substituting this into equation (4.5) gives

$$\begin{aligned}
\lambda u_{ij} &= \frac{1}{2}(\sigma_{ij} + \lambda u_{ii} + \lambda u_{jj}) \\
&= \frac{1}{2}\left(\sigma_{ij} - \frac{1}{m}\sum_{k\neq i}\sigma_{ik} - \frac{1}{m}\sum_{k\neq j}\sigma_{jk} + \frac{1}{m^2}\sum_{k=1}^{m}\sum_{l\neq k}\sigma_{kl}\right)
\end{aligned}$$

for $i \neq j$. It may be checked that if we rescale u_{ij} by setting $\lambda = 1/2$, then we have

$$\sum_{i=1}^{m} \sum_{j \neq i} \sigma_{ij}(Z_i - \overline{Z})(Z_j - \overline{Z}) = \sum_{i=1}^{m} \sum_{j=1}^{m} u_{ij} Z_i Z_j$$

In other words, the corresponding choice of weights w_{ij} in Moran's I statistic is simply $w_{ij} = \sigma_{ij}$ for $i \neq j$, and $w_{ii} = 0$. □

REFERENCES

[1] R. H. Bamberger and M. J. Smith, *A filter bank for the directional decomposition of images: Theory and design*, Signal Processing, IEEE Transactions on, 40 (1992), pp. 882–893.

[2] R. P. Barry and R. Kelley Pace, *Kriging with large data sets using sparse matrix techniques*, Communications in Statistics-Simulation and Computation, 26 (1997), pp. 619–629.

[3] R. C. Bradley, *A caution on mixing conditions for random fields*, Statistics & Probability Letters, 8 (1989), pp. 489–491.

[4] ——, *Introduction to strong mixing conditions*, Kendrick Press, 2007.

[5] R. C. Bradley et al., *Equivalent mixing conditions for random fields*, The Annals of Probability, 21 (1993), pp. 1921–1926.

[6] ——, *Basic properties of strong mixing conditions. a survey and some open questions*, Probability surveys, 2 (2005), pp. 107–144.

[7] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, The Annals of Mathematical Statistics, (1952), pp. 493–507.

[8] N. Cressie and G. Johannesson, *Fixed rank kriging for very large spatial data sets*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70 (2008), pp. 209–226.

[9] N. A. C. Cressie, *Statistics for spatial data, revised edition*, 1993.

[10] B. M. Davis and L. E. Borgman, *A note on the asymptotic distribution of the sample variogram*, Mathematical Geology, 14 (1982), pp. 189–193.

[11] A. Deitmar and S. Echterhoff, *Principles of harmonic analysis*, Springer, 2014.

[12] M. N. Do and M. Vetterli, *The contourlet transform: an efficient directional multiresolution image representation*, Image Processing, IEEE Transactions on, 14 (2005), pp. 2091–2106.

[13] J. Duda, *Asymmetric numeral systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding*, 2013.

[14] P. H. Garcia-Soidán, M. Febrero-Bande, and W. González-Manteiga, *Non-parametric kernel estimation of an isotropic variogram*, Journal of Statistical Planning and Inference, 121 (2004), pp. 65–92.

[15] H. Gebelein, *Das statistische problem der korrelation als variations-und eigen-wertproblem und sein zusammenhang mit der ausgleichsrechnung*, ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik, 21 (1941), pp. 364–379.

[16] A. Gelfand, P. J. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*, CRC Press, 2010.

[17] P. Hall, N. I. Fisher, and B. Hoffmann, *On the nonparametric estimation of covariance functions*, The Annals of Statistics, (1994), pp. 2115–2134.

[18] P. Hall and P. Patil, *Properties of nonparametric estimators of autocovariance for stationary random fields*, Probability Theory and Related Fields, 99 (1994), pp. 399–424.

[19] L. Hartman and O. Hössjer, *Fast kriging of large data sets with gaussian markov random fields*, Computational Statistics & Data Analysis, 52 (2008), pp. 2331–2349.

[20] W. Jarczyk and P. Volkmann, *On functional equations in connection with the absolute value of additive functions*, Series Mathematicae Catoviciensis et Debreceniensis, 32 (2010).

[21] S. Johansen, *An application of extreme point methods to the representation of infinitely divisible distributions*, Probability Theory and Related Fields, 5 (1966), pp. 304–316.

[22] O. Kallenberg, *Foundations of modern probability*, Springer Science & Business Media, 2006.

[23] A. Kolmogorov and Y. A. Rozanov, *On strong mixing conditions for stationary Gaussian processes*, Theory of Probability & Its Applications, 5 (1960), pp. 204–208.

[24] D. G. Krige, *A statistical approach to some basic mine valuation problems on the witwatersrand*, Journal of Chemical, Metallurgical, and Mining Society of South Africa, 52 (1951), pp. 119–139.

[25] ——, *A review of the development of geostatistics in South Africa*, in Advanced Geostatistics in the Mining Industry, vol. 24 of NATO Advanced Study Institutes Series, 1976, pp. 279–293.

[26] S. N. Lahiri, *On inconsistency of estimators based on spatial data under infill asymptotics*, Sankhyā: The Indian Journal of Statistics, Series A, (1996), pp. 403–417.

[27] S. N. Lahiri, Y. Lee, and N. Cressie, *On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters*, Journal of Statistical Planning and Inference, 103 (2002), pp. 65–85.

[28] H. O. Lancaster, *Some properties of the bivariate normal distribution considered in the form of a contingency table*, Biometrika, (1957), pp. 289–292.

[29] E. Le Pennec and S. Mallat, *Sparse geometric image representations with bandelets*, Image Processing, IEEE Transactions on, 14 (2005), pp. 423–438.

[30] G. Mathéron, *Principles of geostatistics*, Economic Geology, 58 (1963), pp. 1246–1266.

[31] ——, *The Theory of Regionalized Variables and Its Applications*, no. 5 in Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, École Nationale Supérieure des Mines de Paris, 1971.

[32] N. Memarsadeghi, V. C. Raykar, R. Duraiswami, and D. M. Mount, *Efficient kriging via fast matrix-vector products*, in Aerospace Conference, 2008 IEEE, IEEE, 2008, pp. 1–7.

[33] A. Rényi, *New version of the probabilistic generalization of the large sieve*, Acta Mathematica Hungarica, 10 (1959), pp. 217–226.

[34] W. Rudin, *Real and complex analysis*, Tata McGraw-Hill Education, 1987.

[35] ——, *Fourier analysis on groups*, no. 12, John Wiley & Sons, 1990.

[36] O. Schabenberger and C. A. Gotway, *Statistical Methods for Spatial Data Analysis*, Chapman and Hall/CRC, 2004.

[37] I. J. Schoenberg, *Metric spaces and completely monotone functions*, Annals of Mathematics, (1938), pp. 811–841.

[38] A. Shapiro and J. Botha, *Variogram fitting with a general class of conditionally nonnegative definite functions*, Computational Statistics & Data Analysis, 11 (1991), pp. 87–96.

[39] G. J. Sullivan, *Efficient scalar quantization of exponential and Laplacian random variables*, Information Theory, IEEE Transactions on, 42 (1996), pp. 1365–1374.

[40] M. B. Wakin, J. K. Romberg, H. Choi, and R. G. Baraniuk, *Geometric methods for wavelet-based image compression*, in Optical Science and Technology, SPIE's 48th Annual Meeting, International Society for Optics and Photonics, 2003, pp. 507–520.

[41] Y. Yu, *On the maximal correlation coefficient*, Statistics & Probability Letters, 78 (2008), pp. 1072–1075.